Notes on the combinatorial fundamentals of algebra*

Darij Grinberg

January 10, 2019
(with minor corrections August 25, 2020)†

Abstract. This is a detailed survey – with rigorous and self-contained proofs – of some of the basics of elementary combinatorics and algebra, including the properties of finite sums, binomial coefficients, permutations and determinants. It is entirely expository (and written to a large extent as a repository for folklore proofs); no new results (and few, if any, new proofs) appear.

Contents

*old title: PRIMES 2015 reading project: problems and solutions
†The numbering in this version is compatible with that in the version of 10 January 2019.

1. Introduction

These notes are a detailed introduction to some of the basic objects of combinatorics and algebra: finite sums, binomial coefficients, permutations and determinants (from a combinatorial viewpoint – no linear algebra is presumed). To a lesser extent, modular arithmetic and recurrent integer sequences are treated as well. The reader is assumed to be proficient in high-school mathematics, and mature enough to understand nontrivial mathematical proofs. Familiarity with "contest mathematics" is also useful.

One feature of these notes is their focus on rigorous and detailed proofs. Indeed, so extensive are the details that a reader with experience in mathematics will probably be able to skip whole paragraphs of proof without losing the thread. (As a consequence of this amount of detail, the notes contain far less material than might be expected from their length.) Rigorous proofs mean that (with some minor exceptions) no "handwaving" is used; all relevant objects are defined in mathematical (usually set-theoretical) language, and are manipulated in logically well-defined ways. (In particular, some things that are commonly taken for granted in the literature – e.g., the fact that the sum of n numbers is well-defined without specifying in what order they are being added – are unpacked and proven in a rigorous way.)

These notes are split into several chapters:

- Chapter 1 collects some basic facts and notations that are used in later chapters. This chapter is **not** meant to be read first; it is best consulted when needed.

- Chapter 2 is an in-depth look at mathematical induction (in various forms, including strong and two-sided induction) and several of its applications (including basic modular arithmetic, division with remainder, Bezout's theorem, some properties of recurrent sequences, the well-definedness of compositions of n maps and sums of n numbers, and various properties thereof).

- Chapter 3 surveys binomial coefficients and their basic properties. Unlike most texts on combinatorics, our treatment of binomial coefficients leans to the algebraic side, relying mostly on computation and manipulations of sums; but some basics of counting are included.

- Chapter 4 treats some more properties of Fibonacci-like sequences, including explicit formulas (à la Binet) for two-term recursions of the form $x_n = ax_{n-1} + bx_{n-2}$.

- Chapter 5 is concerned with permutations of finite sets. The coverage is heavily influenced by the needs of the next chapter (on determinants); thus, a great role is played by transpositions and the inversions of a permutation.

- Chapter 6 is a comprehensive introduction to determinants of square matrices over a commutative ring[1], from an elementary point of view. This is probably the most unique feature of these notes: I define determinants using Leibniz's formula (i.e., as sums over permutations) and prove all their properties (Laplace expansion in one or several rows; the Cauchy-Binet, Desnanot-Jacobi and Plücker identities; the Vandermonde and Cauchy determinants; and several more) from this vantage point, thus treating them as an elementary object unmoored from its linear-algebraic origins and applications. No use is made of modules (or vector spaces), exterior powers, eigenvalues, or of the "universal coefficients" trick[2]. (This means that all proofs are done through combinatorics and manipulation of sums – a rather restrictive requirement!) This is a conscious and (to a large extent) aesthetic choice on my part, and I do **not** consider it the best way to learn about determinants; but I do regard it as a road worth charting, and these notes are my attempt at doing so.

The notes include numerous exercises of varying difficulty, many of them solved. The reader should treat exercises and theorems (and propositions, lemmas and corollaries) as interchangeable to some extent; it is perfectly reasonable to read the solution of an exercise, or conversely, to prove a theorem on one's own instead of reading its proof. The reader's experience will be the strongest determinant of their success in solving the exercises independently.

I have not meant these notes to be a textbook on any particular subject. For one thing, their content does not map to any of the standard university courses, but rather straddles various subjects:

- Much of Chapter 3 (on binomial coefficients) and Chapter 5 (on permutations) is seen in a typical combinatorics class; but my focus is more on the algebraic side and not so much on the combinatorics.

- Chapter 6 studies determinants far beyond what a usual class on linear algebra would do; but it does not include any of the other topics that a linear algebra class usually covers (such as row reduction, vector spaces, linear maps, eigenvectors, tensors or bilinear forms).

- Being devoted to mathematical induction, Chapter 2 appears to cover the same ground as a typical "introduction to proofs" textbook or class (or at least one of its main topics). In reality, however, it complements rather than competes with most "introduction to proofs" texts I have seen; the examples I give are (with a few exceptions) nonstandard, and the focus different.

[1]The notion of a commutative ring is defined (and illustrated with several examples) in Section 6.1, but I don't delve deeper into abstract algebra.

[2]This refers to the standard trick used for proving determinant identities (and other polynomial identities), in which one first replaces the entries of a matrix (or, more generally, the variables appearing in the identity) by indeterminates, then uses the "genericity" of these indeterminates (e.g., to invert the matrix, or to divide by an expression that could otherwise be 0), and finally substitutes the old variables back for the indeterminates.

- While the notions of rings and groups are defined in Chapter 6, I cannot claim to really be doing any abstract algebra: I am merely working *in* rings (i.e., doing computations with elements of rings or with matrices over rings), rather than working *with* rings. Nevertheless, Chapter 6 might help familiarize the reader with these concepts, facilitating proper learning of abstract algebra later on.

All in all, these notes are probably more useful as a repository of detailed proofs than as a textbook to be read cover-to-cover. Indeed, one of my motives in writing them was to have a reference for certain folklore results – one in which these results are proved elementary and without appeal to the reader's problem-solving acumen.

These notes began as worksheets for the PRIMES reading project I have mentored in 2015; they have since been greatly expanded with new material (some of it originally written for my combinatorics classes, some in response to math.stackexchange questions).

The notes are in flux, and probably have their share of misprints. I thank Anya Zhang and Karthik Karnik (the two students taking part in the 2015 PRIMES project) for finding some errors. Thanks also to the PRIMES project at MIT, which gave the impetus for the writing of this notes; and to George Lusztig for the sponsorship of my mentoring position in this project.

1.1. Prerequisites

Let me first discuss the prerequisites for a reader of these notes. At the current moment, I assume that the reader

- has a good grasp on basic school-level mathematics (integers, rational numbers, etc.);
- has some experience with proofs (mathematical induction, proof by contradiction, the concept of "WLOG", etc.) and mathematical notation (functions, subscripts, cases, what it means for an object to be "well-defined", etc.)[3];

[3] A great introduction into these matters (and many others!) is the free book [LeLeMe16] by Lehman, Leighton and Meyer. (**Practical note:** As of 2018, this book is still undergoing frequent revisions; thus, the version I am citing below might be outdated by the time you are reading this. I therefore suggest searching for possibly newer versions on the internet. Unfortunately, you will also find many older versions, often as the first google hits. Try searching for the title of the book along with the current year to find something up-to-date.)

Another introduction to proofs and mathematical workmanship is Day's [Day16] (but beware that the definition of polynomials in [Day16, Chapter 5] is the wrong one for our purposes). Two others are Hammack's [Hammac15] and Doud's and Nielsen's [DouNie19]. Yet another is Newstead's [Newste19] (currently a work in progress, but promising to become one of the most interesting and sophisticated texts of this kind). There are also several books on this subject; an especially popular one is Velleman's [Vellem06].

- knows what a polynomial is (at least over $\mathbb{Z}$ and $\mathbb{Q}$) and how polynomials differ from polynomial functions[4];
- is somewhat familiar with the summation sign ($\sum$) and the product sign ($\prod$) and knows how to transform them (e.g., interchanging summations, and substituting the index)[5];
- has some familiarity with matrices (i.e., knows how to add and to multiply them)[6].

Probably a few more requirements creep in at certain points of the notes, which I have overlooked. Some examples and remarks rely on additional knowledge (such as analysis, graph theory, abstract algebra); however, these can be skipped.

1.2. Notations

- In the following, we use $\mathbb{N}$ to denote the set $\{0, 1, 2, \ldots\}$. (Be warned that some other authors use the letter $\mathbb{N}$ for $\{1, 2, 3, \ldots\}$ instead.)
- We let $\mathbb{Q}$ denote the set of all rational numbers; we let $\mathbb{R}$ be the set of all real numbers; we let $\mathbb{C}$ be the set of all complex numbers[7].
- If X and Y are two sets, then we shall use the notation "$X \to Y,\ x \mapsto E$" (where x is some symbol which has no specific meaning in the current context, and where E is some expression which usually involves x) for "the map from X to Y which sends every $x \in X$ to E".

  For example, "$\mathbb{N} \to \mathbb{N},\ x \mapsto x^2 + x + 6$" means the map from $\mathbb{N}$ to $\mathbb{N}$ which sends every $x \in \mathbb{N}$ to $x^2 + x + 6$.

  For another example, "$\mathbb{N} \to \mathbb{Q},\ x \mapsto \dfrac{x}{1+x}$" denotes the map from $\mathbb{N}$ to $\mathbb{Q}$ which sends every $x \in \mathbb{N}$ to $\dfrac{x}{1+x}$. [8]

[4] This is used only in a few sections and exercises, so it is not an unalienable requirement. See Section 1.5 below for a quick survey of polynomials, and for references to sources in which precise definitions can be found.

[5] See Section 1.4 below for a quick overview of the notations that we will need.

[6] See, e.g., [Grinbe16b, Chapter 2] or any textbook on linear algebra for an introduction.

[7] See [Swanso18, Section 3.9] or [AmaEsc05, Section I.11] for a quick introduction to complex numbers. We will rarely use complex numbers. Most of the time we use them, you can instead use real numbers.

[8] A word of warning: Of course, the notation "$X \to Y,\ x \mapsto E$" does not always make sense; indeed, the map that it stands for might sometimes not exist. For instance, the notation "$\mathbb{N} \to \mathbb{Q},\ x \mapsto \dfrac{x}{1-x}$" does not actually define a map, because the map that it is supposed to define (i.e., the map from $\mathbb{N}$ to $\mathbb{Q}$ which sends every $x \in \mathbb{N}$ to $\dfrac{x}{1-x}$) does not exist (since $\dfrac{x}{1-x}$ is not defined for $x = 1$). For another example, the notation "$\mathbb{N} \to \mathbb{Z},\ x \mapsto \dfrac{x}{1+x}$" does not define

- If S is a set, then the *powerset* of S means the set of all subsets of S. This powerset will be denoted by $\mathcal{P}(S)$. For example, the powerset of $\{1,2\}$ is $\mathcal{P}(\{1,2\}) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$.
- The letter i will **not** denote the imaginary unit $\sqrt{-1}$ (except when we explicitly say so).

Further notations will be defined whenever they arise for the first time.

1.3. Injectivity, surjectivity, bijectivity

In this section[9], we recall some basic properties of maps – specifically, what it means for a map to be injective, surjective and bijective. We begin by recalling basic definitions:

- The words "map", "mapping", "function", "transformation" and "operator" are synonyms in mathematics.[10]
- A map $f : X \to Y$ between two sets X and Y is said to be *injective* if it has the following property:
  - If x_1 and x_2 are two elements of X satisfying $f(x_1) = f(x_2)$, then $x_1 = x_2$. (In words: If two elements of X are sent to one and the same element of Y by f, then these two elements of X must have been equal in the first place. In other words: An element of X is uniquely determined by its image under f.)

  Injective maps are often called "one-to-one maps" or "injections".

  For example:

  - The map $\mathbb{Z} \to \mathbb{Z},\ x \mapsto 2x$ (this is the map that sends each integer x to $2x$) is injective, because if x_1 and x_2 are two integers satisfying $2x_1 = 2x_2$, then $x_1 = x_2$.
  - The map $\mathbb{Z} \to \mathbb{Z},\ x \mapsto x^2$ (this is the map that sends each integer x to x^2) is **not** injective, because if x_1 and x_2 are two integers satisfying $x_1^2 = x_2^2$, then we do not necessarily have $x_1 = x_2$. (For example, if $x_1 = -1$ and $x_2 = 1$, then $x_1^2 = x_2^2$ but not $x_1 = x_2$.)

a map, because the map that it is supposed to define (i.e., the map from $\mathbb{N}$ to $\mathbb{Z}$ which sends every $x \in \mathbb{N}$ to $\dfrac{x}{1+x}$) does not exist (for $x = 2$, we have $\dfrac{x}{1+x} = \dfrac{2}{1+2} \notin \mathbb{Z}$, which shows that a map from $\mathbb{N}$ to $\mathbb{Z}$ cannot send this x to this $\dfrac{x}{1+x}$). Thus, when defining a map from X to Y (using whatever notation), do not forget to check that it is well-defined (i.e., that your definition specifies precisely one image for each $x \in X$, and that these images all lie in Y). In many cases, this is obvious or very easy to check (I will usually not even mention this check), but in some cases, this is a difficult task.

[9]a significant part of which is copied from [Grinbe16b, §3.21]

[10]That said, mathematicians often show some nuance by using one of them and not the other. However, we do not need to concern ourselves with this here.

- A map $f : X \to Y$ between two sets X and Y is said to be *surjective* if it has the following property:

  - For each $y \in Y$, there exists some $x \in X$ satisfying $f(x) = y$. (In words: Each element of Y is an image of some element of X under f.)

  Surjective maps are often called "onto maps" or "surjections".

  For example:

  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x + 1$ (this is the map that sends each integer x to $x + 1$) is surjective, because each integer y has some integer satisfying $x + 1 = y$ (namely, $x = y - 1$).
  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto 2x$ (this is the map that sends each integer x to $2x$) is **not** surjective, because not each integer y has some integer x satisfying $2x = y$. (For instance, $y = 1$ has no such x, since y is odd.)
  - The map $\{1, 2, 3, 4\} \to \{1, 2, 3, 4, 5\}$, $x \mapsto x$ (this is the map sending each x to x) is **not** surjective, because not each $y \in \{1, 2, 3, 4, 5\}$ has some $x \in \{1, 2, 3, 4\}$ satisfying $x = y$. (Namely, $y = 5$ has no such x.)

- A map $f : X \to Y$ between two sets X and Y is said to be *bijective* if it is both injective and surjective. Bijective maps are often called "one-to-one correspondences" or "bijections".

  For example:

  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x + 1$ is bijective, since it is both injective and surjective.
  - The map $\{1, 2, 3, 4\} \to \{1, 2, 3, 4, 5\}$, $x \mapsto x$ is **not** bijective, since it is not surjective. (However, it is injective.)
  - The map $\mathbb{Z} \to \mathbb{N}$, $x \mapsto |x|$ is **not** bijective, since it is not injective. (However, it is surjective.)
  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x^2$ is **not** bijective, since it is not injective. (It also is not surjective.)

- If X is a set, then id_X denotes the map from X to X that sends each $x \in X$ to x itself. (In words: id_X denotes the map which sends each element of X to itself.) The map id_X is often called the *identity map on* X, and often denoted by id (when X is clear from the context or irrelevant). The identity map id_X is always bijective.

- If $f : X \to Y$ and $g : Y \to Z$ are two maps, then the *composition* $g \circ f$ of the maps g and f is defined to be the map from X to Z that sends each $x \in X$ to $g(f(x))$. (In words: The composition $g \circ f$ is the map from X to Z that applies the map f **first** and **then** applies the map g.) You might find it confusing that this map is denoted by $g \circ f$ (rather than $f \circ g$), given that it proceeds by applying f first and g last; however, this has its reasons:

It satisfies $(g \circ f)(x) = g(f(x))$. Had we denoted it by $f \circ g$ instead, this equality would instead become $(f \circ g)(x) = g(f(x))$, which would be even more confusing.

- If $f : X \to Y$ is a map between two sets X and Y, then an *inverse* of f means a map $g : Y \to X$ satisfying $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. (In words, the condition "$f \circ g = \mathrm{id}_Y$" means "if you start with some element $y \in Y$, then apply g, then apply f, then you get y back", or equivalently "the map f undoes the map g". Similarly, the condition "$g \circ f = \mathrm{id}_X$" means "if you start with some element $x \in X$, then apply f, then apply g, then you get x back", or equivalently "the map g undoes the map f". Thus, an inverse of f means a map $g : Y \to X$ that both undoes and is undone by f.)

  The map $f : X \to Y$ is said to be *invertible* if and only if an inverse of f exists. If an inverse of f exists, then it is unique[11], and thus is called *the inverse of* f, and is denoted by f^{-1}.

  For example:

  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x + 1$ is invertible, and its inverse is $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x - 1$.
  - The map $\mathbb{Q} \setminus \{1\} \to \mathbb{Q} \setminus \{0\}$, $x \mapsto \dfrac{1}{1-x}$ is invertible, and its inverse is the map $\mathbb{Q} \setminus \{0\} \to \mathbb{Q} \setminus \{1\}$, $x \mapsto 1 - \dfrac{1}{x}$.

- If $f : X \to Y$ is a map between two sets X and Y, then the following notations will be used:

  - For any subset U of X, we let $f(U)$ be the subset $\{f(u) \mid u \in U\}$ of Y. This set $f(U)$ is called the *image* of U under f. This should not be confused with the image $f(x)$ of a single element $x \in X$ under f.

---

[11] *Proof.* Let g_1 and g_2 be two inverses of f. We shall show that $g_1 = g_2$.

We know that g_1 is an inverse of f. In other words, g_1 is a map $Y \to X$ satisfying $f \circ g_1 = \mathrm{id}_Y$ and $g_1 \circ f = \mathrm{id}_X$.

We know that g_2 is an inverse of f. In other words, g_2 is a map $Y \to X$ satisfying $f \circ g_2 = \mathrm{id}_Y$ and $g_2 \circ f = \mathrm{id}_X$.

A well-known fact (known as *associativity of map composition*, and stated explicitly as Proposition 2.82 below) says that if X, Y, Z and W are four sets, and if $c : X \to Y$, $b : Y \to Z$ and $a : Z \to W$ are three maps, then

$$(a \circ b) \circ c = a \circ (b \circ c).$$

Applying this fact to Y, X, Y, X, g_1, f and g_2 instead of X, Y, Z, W, c, b and a, we obtain $(g_2 \circ f) \circ g_1 = g_2 \circ (f \circ g_1)$.

Hence, $g_2 \circ (f \circ g_1) = \underbrace{(g_2 \circ f)}_{=\mathrm{id}_X} \circ g_1 = \mathrm{id}_X \circ g_1 = g_1$. Comparing this with $g_2 \circ \underbrace{(f \circ g_1)}_{=\mathrm{id}_Y} = g_2 \circ \mathrm{id}_Y = g_2$, we obtain $g_1 = g_2$.

Now, forget that we fixed g_1 and g_2. We thus have shown that if g_1 and g_2 are two inverses of f, then $g_1 = g_2$. In other words, any two inverses of f must be equal. In other words, if an inverse of f exists, then it is unique.

Note that the map $f : X \to Y$ is surjective if and only if $Y = f(X)$. (This is easily seen to be a restatement of the definition of "surjective".)

- For any subset V of Y, we let $f^{-1}(V)$ be the subset $\{u \in X \mid f(u) \in V\}$ of X. This set $f^{-1}(V)$ is called the *preimage* of V under f. This should not be confused with the image $f^{-1}(y)$ of a single element $y \in Y$ under the inverse f^{-1} of f (when this inverse exists).

  (Note that in general, $f\left(f^{-1}(V)\right) \neq V$ and $f^{-1}(f(U)) \neq U$. However, $f\left(f^{-1}(V)\right) \subseteq V$ and $U \subseteq f^{-1}(f(U))$.)

- For any subset U of X, we let $f\mid_U$ be the map from U to Y which sends each $u \in U$ to $f(u) \in Y$. This map $f\mid_U$ is called the *restriction* of f to the subset U.

The following facts are fundamental:

Theorem 1.1. A map $f : X \to Y$ is invertible if and only if it is bijective.

Theorem 1.2. Let U and V be two finite sets. Then, $|U| = |V|$ if and only if there exists a bijective map $f : U \to V$.

Theorem 1.2 holds even if the sets U and V are infinite, but to make sense of this we would need to define the size of an infinite set, which is a much subtler issue than the size of a finite set. We will only need Theorem 1.2 for finite sets.

Let us state some more well-known and basic properties of maps between finite sets:

Lemma 1.3. Let U and V be two finite sets. Let $f : U \to V$ be a map.

(a) We have $|f(S)| \leq |S|$ for each subset S of U.

(b) Assume that $|f(U)| \geq |U|$. Then, the map f is injective.

(c) If f is injective, then $|f(S)| = |S|$ for each subset S of U.

Lemma 1.4. Let U and V be two finite sets such that $|U| \leq |V|$. Let $f : U \to V$ be a map. Then, we have the following logical equivalence:

$$(f \text{ is surjective}) \iff (f \text{ is bijective}).$$

Lemma 1.5. Let U and V be two finite sets such that $|U| \geq |V|$. Let $f : U \to V$ be a map. Then, we have the following logical equivalence:

$$(f \text{ is injective}) \iff (f \text{ is bijective}).$$

Exercise 1.1. Prove Lemma 1.3, Lemma 1.4 and Lemma 1.5.

Let us make one additional observation about maps:

Remark 1.6. Composition of maps is associative: If X, Y, Z and W are three sets, and if $c : X \to Y$, $b : Y \to Z$ and $a : Z \to W$ are three maps, then $(a \circ b) \circ c = a \circ (b \circ c)$. (This shall be proven in Proposition 2.82 below.)

In Section 2.13, we shall prove a more general fact: If $X_1, X_2, \ldots, X_{k+1}$ are $k+1$ sets for some $k \in \mathbb{N}$, and if $f_i : X_i \to X_{i+1}$ is a map for each $i \in \{1, 2, \ldots, k\}$, then the composition $f_k \circ f_{k-1} \circ \cdots \circ f_1$ of all k maps $f_1, f_2, \ldots, f_k$ is a well-defined map from X_1 to X_{k+1}, which sends each element $x \in X_1$ to $f_k(f_{k-1}(f_{k-2}(\cdots(f_2(f_1(x)))\cdots)))$ (in other words, which transforms each element $x \in X_1$ by first applying f_1, then applying f_2, then applying f_3, and so on); this composition $f_k \circ f_{k-1} \circ \cdots \circ f_1$ can also be written as $f_k \circ (f_{k-1} \circ (f_{k-2} \circ (\cdots \circ (f_2 \circ f_1) \cdots)))$ or as $(((\cdots (f_k \circ f_{k-1}) \circ \cdots) \circ f_3) \circ f_2) \circ f_1$. An important particular case is when $k = 0$; in this case, $f_k \circ f_{k-1} \circ \cdots \circ f_1$ is a composition of 0 maps. It is defined to be id_{X_1} (the identity map of the set X_1), and it is called the "empty composition of maps $X_1 \to X_1$". (The logic behind this definition is that the composition $f_k \circ f_{k-1} \circ \cdots \circ f_1$ should transform each element $x \in X_1$ by first applying f_1, then applying f_2, then applying f_3, and so on; but for $k = 0$, there are no maps to apply, and so x just remains unchanged.)

1.4. Sums and products: a synopsis

In this section, I will recall the definitions of the $\sum$ and $\prod$ signs and collect some of their basic properties (without proofs). When I say "recall", I am implying that the reader has at least some prior acquaintance (and, ideally, experience) with these signs; for a first introduction, this section is probably too brief and too abstract. Ideally, you should use this section to familiarize yourself with my (sometimes idiosyncratic) notations.

Throughout Section 1.4, we let $\mathbb{A}$ be one of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.

1.4.1. Definition of $\sum$

Let us first define the $\sum$ sign. There are actually several (slightly different, but still closely related) notations involving the $\sum$ sign; let us define the most important of them:

- If S is a finite set, and if a_s is an element of $\mathbb{A}$ for each $s \in S$, then $\sum_{s \in S} a_s$ denotes the sum of all of these elements a_s. Formally, this sum is defined by recursion on $|S|$, as follows:
  - If $|S| = 0$, then $\sum_{s \in S} a_s$ is defined to be 0.

– Let $n \in \mathbb{N}$. Assume that we have defined $\sum_{s \in S} a_s$ for every finite set S with $|S| = n$ (and every choice of elements a_s of $\mathbb{A}$). Now, if S is a finite set with $|S| = n + 1$ (and if $a_s \in \mathbb{A}$ are chosen for all $s \in S$), then $\sum_{s \in S} a_s$ is defined by picking any $t \in S$ [12] and setting

$$\sum_{s \in S} a_s = a_t + \sum_{s \in S \setminus \{t\}} a_s. \tag{1}$$

It is not immediately clear why this definition is legitimate: The right hand side of (1) is defined using a choice of t, but we want our value of $\sum_{s \in S} a_s$ to depend only on S and on the a_s (not on some arbitrarily chosen $t \in S$). However, it is possible to prove that the right hand side of (1) is actually independent of t (that is, any two choices of t will lead to the same result). See Section 2.14 below (and Theorem 2.118 **(a)** in particular) for the proof of this fact.

Examples:

– If $S = \{4, 7, 9\}$ and $a_s = \frac{1}{s^2}$ for every $s \in S$, then $\sum_{s \in S} a_s = a_4 + a_7 + a_9 = \frac{1}{4^2} + \frac{1}{7^2} + \frac{1}{9^2} = \frac{6049}{63504}$.

– If $S = \{1, 2, \ldots, n\}$ (for some $n \in \mathbb{N}$) and $a_s = s^2$ for every $s \in S$, then $\sum_{s \in S} a_s = \sum_{s \in S} s^2 = 1^2 + 2^2 + \cdots + n^2$. (There is a formula saying that the right hand side of this equality is $\frac{1}{6} n (2n + 1)(n + 1)$.)

– If $S = \varnothing$, then $\sum_{s \in S} a_s = 0$ (since $|S| = 0$).

Remarks:

– The sum $\sum_{s \in S} a_s$ is usually pronounced "sum of the a_s over all $s \in S$" or "sum of the a_s with s ranging over S" or "sum of the a_s with s running through all elements of S". The letter "s" in the sum is called the "summation index"[13], and its exact choice is immaterial (for example, you can rewrite $\sum_{s \in S} a_s$ as $\sum_{t \in S} a_t$ or as $\sum_{\Phi \in S} a_\Phi$ or as $\sum_{\spadesuit \in S} a_\spadesuit$), as long as it does not already have a different meaning outside of the sum[14]. (Ultimately,

[12] This is possible, because S is nonempty (in fact, $|S| = n + 1 > n \geq 0$).

[13] The plural of the word "index" here is "indices", not "indexes".

[14] If it already has a different meaning, then it must not be used as a summation index! For example, you must not write "every $n \in \mathbb{N}$ satisfies $\sum_{n \in \{0,1,\ldots,n\}} n = \frac{n(n+1)}{2}$", because here the summation index n clashes with a different meaning of the letter n.

a summation index is the same kind of placeholder variable as the "s" in the statement "for all $s \in S$, we have $a_s + 2a_s = 3a_s$", or as a loop variable in a for-loop in programming.) The sign $\sum$ itself is called "the summation sign" or "the $\sum$ sign". The numbers a_s are called the *addends* (or *summands*) of the sum $\sum\limits_{s\in S} a_s$. More precisely, for any given $t \in S$, we can refer to the number a_t as the "addend corresponding to the index t" (or as the "addend for $s = t$", or as the "addend for t") of the sum $\sum\limits_{s\in S} a_s$.

- When the set S is empty, the sum $\sum\limits_{s\in S} a_s$ is called an *empty sum*. Our definition implies that any empty sum is 0. This convention is used throughout mathematics, except in rare occasions where a slightly subtler version of it is used[15]. Ignore anyone who tells you that empty sums are undefined!

- The summation index does not always have to be a single letter. For instance, if S is a set of pairs, then we can write $\sum\limits_{(x,y)\in S} a_{(x,y)}$ (meaning the same as $\sum\limits_{s\in S} a_s$). Here is an example of this notation:

$$\sum_{(x,y)\in\{1,2,3\}^2} \frac{x}{y} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{2}{1} + \frac{2}{2} + \frac{2}{3} + \frac{3}{1} + \frac{3}{2} + \frac{3}{3}$$

  (here, we are using the notation $\sum\limits_{(x,y)\in S} a_{(x,y)}$ with $S = \{1,2,3\}^2$ and $a_{(x,y)} = \frac{x}{y}$). Note that we could not have rewritten this sum in the form $\sum\limits_{s\in S} a_s$ with a single-letter variable s without introducing an extra notation such as $a_{(x,y)}$ for the quotients $\frac{x}{y}$.

- Mathematicians don't seem to have reached an agreement on the operator precedence of the $\sum$ sign. By this I mean the following question:

[15]Do not worry about this subtler version for the time being. If you really want to know what it is: Our above definition is tailored to the cases when the a_s are numbers (i.e., elements of one of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$). In more advanced settings, one tends to take sums of the form $\sum\limits_{s\in S} a_s$ where the a_s are not numbers but (for example) elements of a commutative ring $\mathbb{K}$. (See Definition 6.2 for the definition of a commutative ring.) In such cases, one wants the sum $\sum\limits_{s\in S} a_s$ for an empty set S to be not the integer 0, but the zero of the commutative ring $\mathbb{K}$ (which is sometimes distinct from the integer 0). This has the slightly confusing consequence that the meaning of the sum $\sum\limits_{s\in S} a_s$ for an empty set S depends on what ring $\mathbb{K}$ the a_s belong to, even if (for an empty set S) there are no a_s to begin with! But in practice, the choice of $\mathbb{K}$ is always clear from context, so this is not ambiguous.

A similar caveat applies to the other versions of the $\sum$ sign, as well as to the $\prod$ sign defined further below; I shall not elaborate on it further.

Does $\sum_{s\in S} a_s + b$ (where b is some other element of $\mathbb{A}$) mean $\sum_{s\in S} (a_s + b)$ or $\left(\sum_{s\in S} a_s\right) + b$? In my experience, the second interpretation (i.e., reading it as $\left(\sum_{s\in S} a_s\right) + b$) is more widespread, and this is the interpretation that I will follow. Nevertheless, be on the watch for possible misunderstandings, as someone might be using the first interpretation when you expect it the least![16]

However, the situation is different for products and nested sums. For instance, the expression $\sum_{s\in S} ba_s c$ is understood to mean $\sum_{s\in S} (ba_s c)$, and a nested sum like $\sum_{s\in S}\sum_{t\in T} a_{s,t}$ (where S and T are two sets, and where $a_{s,t}$ is an element of $\mathbb{A}$ for each pair $(s,t) \in S \times T$) is to be read as $\sum_{s\in S}\left(\sum_{t\in T} a_{s,t}\right)$.

- Speaking of nested sums: they mean exactly what they seem to mean. For instance, $\sum_{s\in S}\sum_{t\in T} a_{s,t}$ is what you get if you compute the sum $\sum_{t\in T} a_{s,t}$ for each $s \in S$, and then sum up all of these sums together. In a nested sum $\sum_{s\in S}\sum_{t\in T} a_{s,t}$, the first summation sign ($\sum_{s\in S}$) is called the "outer summation", and the second summation sign ($\sum_{t\in T}$) is called the "inner summation".

- An expression of the form " $\sum_{s\in S} a_s$" (where S is a finite set) is called a *finite sum*.

- We have required the set S to be finite when defining $\sum_{s\in S} a_s$. Of course, this requirement was necessary for our definition, and there is no way to make sense of infinite sums such as $\sum_{s\in\mathbb{Z}} s^2$. However, **some** infinite sums can be made sense of. The simplest case is when the set S might be infinite, but only finitely many among the a_s are nonzero. In this case, we can define $\sum_{s\in S} a_s$ simply by discarding the zero addends and summing the finitely many remaining addends. Other situations in which infinite sums make sense appear in analysis and in topological algebra (e.g., power series).

- The sum $\sum_{s\in S} a_s$ always belongs to $\mathbb{A}$. [17] For instance, a sum of elements of $\mathbb{N}$ belongs to $\mathbb{N}$; a sum of elements of $\mathbb{R}$ belongs to $\mathbb{R}$, and so on.

• A slightly more complicated version of the summation sign is the following: Let S be a finite set, and let $\mathcal{A}(s)$ be a logical statement defined for every

[16] This is similar to the notorious disagreement about whether a/bc means $(a/b)\cdot c$ or $a/(bc)$.

[17] Recall that we have assumed $\mathbb{A}$ to be one of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, and that we have assumed the a_s to belong to $\mathbb{A}$.

$s \in S$ [18]. For example, S can be $\{1, 2, 3, 4\}$, and $\mathcal{A}(s)$ can be the statement "s is even". For each $s \in S$ satisfying $\mathcal{A}(s)$, let a_s be an element of $\mathbb{A}$. Then, the sum $\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s$ is defined by

$$\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s = \sum_{s \in \{t \in S \mid \mathcal{A}(t)\}} a_s.$$

In other words, $\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s$ is the sum of the a_s for all $s \in S$ which satisfy $\mathcal{A}(s)$.

Examples:

– If $S = \{1, 2, 3, 4, 5\}$, then $\sum_{\substack{s \in S; \\ s \text{ is even}}} a_s = a_2 + a_4$. (Of course, $\sum_{\substack{s \in S; \\ s \text{ is even}}} a_s$ is $\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s$ when $\mathcal{A}(s)$ is defined to be the statement "s is even".)

– If $S = \{1, 2, \ldots, n\}$ (for some $n \in \mathbb{N}$) and $a_s = s^2$ for every $s \in S$, then $\sum_{\substack{s \in S; \\ s \text{ is even}}} a_s = a_2 + a_4 + \cdots + a_k$, where k is the largest even number among $1, 2, \ldots, n$ (that is, $k = n$ if n is even, and $k = n - 1$ otherwise).

Remarks:

– The sum $\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s$ is usually pronounced "sum of the a_s over all $s \in S$ satisfying $\mathcal{A}(s)$". The semicolon after "$s \in S$" is often omitted or replaced by a colon or a comma. Many authors often omit the "$s \in S$" part (so they simply write $\sum_{\mathcal{A}(s)} a_s$) when it is clear enough what the S is. (For instance, they would write $\sum_{1 \leq s \leq 5} s^2$ instead of $\sum_{\substack{s \in \mathbb{N}; \\ 1 \leq s \leq 5}} s^2$.)

– The set S needs not be finite in order for $\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s$ to be defined; it suffices that the set $\{t \in S \mid \mathcal{A}(t)\}$ be finite (i.e., that only finitely many $s \in S$ satisfy $\mathcal{A}(s)$).

– The sum $\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s$ is said to be *empty* whenever the set $\{t \in S \mid \mathcal{A}(t)\}$ is empty (i.e., whenever no $s \in S$ satisfies $\mathcal{A}(s)$).

[18] Formally speaking, this means that $\mathcal{A}$ is a map from S to the set of all logical statements. Such a map is called a *predicate*.

- Finally, here is the simplest version of the summation sign: Let u and v be two integers. We agree to understand the set $\{u, u+1, \ldots, v\}$ to be empty when $u > v$. Let a_s be an element of $\mathbb{A}$ for each $s \in \{u, u+1, \ldots, v\}$. Then, $\sum\limits_{s=u}^{v} a_s$ is defined by

$$\sum_{s=u}^{v} a_s = \sum_{s \in \{u, u+1, \ldots, v\}} a_s.$$

Examples:

  - We have $\sum\limits_{s=3}^{8} \frac{1}{s} = \sum\limits_{s \in \{3,4,\ldots,8\}} \frac{1}{s} = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} = \frac{341}{280}$.
  - We have $\sum\limits_{s=3}^{3} \frac{1}{s} = \sum\limits_{s \in \{3\}} \frac{1}{s} = \frac{1}{3}$.
  - We have $\sum\limits_{s=3}^{2} \frac{1}{s} = \sum\limits_{s \in \varnothing} \frac{1}{s} = 0$.

Remarks:

  - The sum $\sum\limits_{s=u}^{v} a_s$ is usually pronounced "sum of the a_s for all s from u to v (inclusive)". It is often written $a_u + a_{u+1} + \cdots + a_v$, but this latter notation has its drawbacks: In order to understand an expression like $a_u + a_{u+1} + \cdots + a_v$, one needs to correctly guess the pattern (which can be unintuitive when the a_s themselves are complicated: for example, it takes a while to find the "moving parts" in the expression $\frac{2 \cdot 7}{3+2} + \frac{3 \cdot 7}{3+3} + \cdots + \frac{7 \cdot 7}{3+7}$, whereas the notation $\sum\limits_{s=2}^{7} \frac{s \cdot 7}{3+s}$ for the same sum is perfectly clear).
  - In the sum $\sum\limits_{s=u}^{v} a_s$, the integer u is called the *lower limit* (of the sum), whereas the integer v is called the *upper limit* (of the sum). The sum is said to *start* (or *begin*) at u and *end* at v.
  - The sum $\sum\limits_{s=u}^{v} a_s$ is said to be *empty* whenever $u > v$. In other words, a sum of the form $\sum\limits_{s=u}^{v} a_s$ is empty whenever it "ends before it has begun". However, a sum which "ends right after it begins" (i.e., a sum $\sum\limits_{s=u}^{v} a_s$ with $u = v$) is not empty; it just has one addend only. (This is unlike integrals, which are 0 whenever their lower and upper limit are equal.)
  - Let me stress once again that a sum $\sum\limits_{s=u}^{v} a_s$ with $u > v$ is empty and equals 0. It does not matter how much greater u is than v. So, for

example, $\sum\limits_{s=1}^{-5} s = 0$. The fact that the upper bound (-5) is much smaller than the lower bound (1) does not mean that you have to subtract rather than add.

Thus we have introduced the main three forms of the summation sign. Some mild variations on them appear in the literature (e.g., there is a slightly awkward notation $\sum\limits_{\substack{s=u;\\ \mathcal{A}(s)}}^{v} a_s$ for $\sum\limits_{\substack{s\in\{u,u+1,\ldots,v\};\\ \mathcal{A}(s)}} a_s$).

1.4.2. Properties of $\sum$

Let me now show some basic properties of summation signs that are important in making them useful:

- **Splitting-off:** Let S be a finite set. Let $t \in S$. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\sum_{s\in S} a_s = a_t + \sum_{s\in S\setminus\{t\}} a_s. \qquad (2)$$
(This is precisely the equality (1) (applied to $n = |S \setminus \{t\}|$), because $|S| = |S \setminus \{t\}| + 1$.) This formula (2) allows us to "split off" an addend from a sum.
Example: If $n \in \mathbb{N}$, then
$$\sum_{s\in\{1,2,\ldots,n+1\}} a_s = a_{n+1} + \sum_{s\in\{1,2,\ldots,n\}} a_s$$
(by (2), applied to $S = \{1, 2, \ldots, n+1\}$ and $t = n+1$), but also
$$\sum_{s\in\{1,2,\ldots,n+1\}} a_s = a_1 + \sum_{s\in\{2,3,\ldots,n+1\}} a_s$$
(by (2), applied to $S = \{1, 2, \ldots, n+1\}$ and $t = 1$).

- **Splitting:** Let S be a finite set. Let X and Y be two subsets of S such that $X \cap Y = \varnothing$ and $X \cup Y = S$. (Equivalently, X and Y are two subsets of S such that each element of S lies in **exactly** one of X and Y.) Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\sum_{s\in S} a_s = \sum_{s\in X} a_s + \sum_{s\in Y} a_s. \qquad (3)$$
(Here, as we explained, $\sum\limits_{s\in X} a_s + \sum\limits_{s\in Y} a_s$ stands for $\left(\sum\limits_{s\in X} a_s\right) + \left(\sum\limits_{s\in Y} a_s\right)$.) The idea behind (3) is that if we want to add a bunch of numbers (the a_s for $s \in S$), we can proceed by splitting it into two "sub-bunches" (one "sub-bunch" consisting of the a_s for $s \in X$, and the other consisting of the a_s for

$s \in Y$), then take the sum of each of these two sub-bunches, and finally add together the two sums. For a rigorous proof of (3), see Theorem 2.130 below.

Examples:

– If $n \in \mathbb{N}$, then

$$\sum_{s\in\{1,2,\ldots,2n\}} a_s = \sum_{s\in\{1,3,\ldots,2n-1\}} a_s + \sum_{s\in\{2,4,\ldots,2n\}} a_s$$

(by (3), applied to $S = \{1,2,\ldots,2n\}$, $X = \{1,3,\ldots,2n-1\}$ and $Y = \{2,4,\ldots,2n\}$).

– If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, then

$$\sum_{s\in\{-m,-m+1,\ldots,n\}} a_s = \sum_{s\in\{-m,-m+1,\ldots,0\}} a_s + \sum_{s\in\{1,2,\ldots,n\}} a_s$$

(by (3), applied to $S = \{-m,-m+1,\ldots,n\}$, $X = \{-m,-m+1,\ldots,0\}$ and $Y = \{1,2,\ldots,n\}$).

– If u, v and w are three integers such that $u-1 \leq v \leq w$, and if a_s is an element of $\mathbb{A}$ for each $s \in \{u,u+1,\ldots,w\}$, then

$$\sum_{s=u}^{w} a_s = \sum_{s=u}^{v} a_s + \sum_{s=v+1}^{w} a_s. \tag{4}$$

This follows from (3), applied to $S = \{u,u+1,\ldots,w\}$, $X = \{u,u+1,\ldots,v\}$ and $Y = \{v+1,v+2,\ldots,w\}$. Notice that the requirement $u-1 \leq v \leq w$ is important; otherwise, the $X \cap Y = \varnothing$ and $X \cup Y = S$ condition would not hold!

- **Splitting using a predicate:** Let S be a finite set. Let $\mathcal{A}(s)$ be a logical statement for each $s \in S$. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s\in S} a_s = \sum_{\substack{s\in S;\\ \mathcal{A}(s)}} a_s + \sum_{\substack{s\in S;\\ \text{not } \mathcal{A}(s)}} a_s \tag{5}$$

(where "not $\mathcal{A}(s)$" means the negation of $\mathcal{A}(s)$). This simply follows from (3), applied to $X = \{s \in S \mid \mathcal{A}(s)\}$ and $Y = \{s \in S \mid \text{not } \mathcal{A}(s)\}$.

Example: If $S \subseteq \mathbb{Z}$, then

$$\sum_{s\in S} a_s = \sum_{\substack{s\in S;\\ s \text{ is even}}} a_s + \sum_{\substack{s\in S;\\ s \text{ is odd}}} a_s$$

(because "s is odd" is the negation of "s is even").

- **Summing equal values:** Let S be a finite set. Let a be an element of $\mathbb{A}$. Then,

$$\sum_{s \in S} a = |S| \cdot a. \qquad (6)$$

[19] In other words, if all addends of a sum are equal to one and the same element a, then the sum is just the number of its addends times a. In particular,

$$\sum_{s \in S} 1 = |S| \cdot 1 = |S| .$$

- **Splitting an addend:** Let S be a finite set. For every $s \in S$, let a_s and b_s be elements of $\mathbb{A}$. Then,

$$\sum_{s \in S} (a_s + b_s) = \sum_{s \in S} a_s + \sum_{s \in S} b_s. \qquad (7)$$

For a rigorous proof of this equality, see Theorem 2.122 below.

Remark: Of course, similar rules hold for other forms of summations: If $\mathcal{A}(s)$ is a logical statement for each $s \in S$, then

$$\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} (a_s + b_s) = \sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s + \sum_{\substack{s \in S; \\ \mathcal{A}(s)}} b_s.$$

If u and v are two integers, then

$$\sum_{s=u}^{v} (a_s + b_s) = \sum_{s=u}^{v} a_s + \sum_{s=u}^{v} b_s. \qquad (8)$$

- **Factoring out:** Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{A}$. Also, let λ be an element of $\mathbb{A}$. Then,

$$\sum_{s \in S} \lambda a_s = \lambda \sum_{s \in S} a_s. \qquad (9)$$

For a rigorous proof of this equality, see Theorem 2.124 below.

Again, similar rules hold for the other types of summation sign.

Remark: Applying (9) to $\lambda = -1$, we obtain

$$\sum_{s \in S} (-a_s) = -\sum_{s \in S} a_s.$$

[19] This is easy to prove by induction on $|S|$.

- **Zeroes sum to zero:** Let S be a finite set. Then,

$$\sum_{s\in S} 0 = 0. \tag{10}$$

That is, any sum of zeroes is zero.

For a rigorous proof of this equality, see Theorem 2.126 below.

Remark: This applies even to infinite sums! Do not be fooled by the infiniteness of a sum: There are no reasonable situations where an infinite sum of zeroes is defined to be anything other than zero. The infinity does not "compensate" for the zero.

- **Dropping zeroes:** Let S be a finite set. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Let T be a subset of S such that every $s \in T$ satisfies $a_s = 0$. Then,

$$\sum_{s\in S} a_s = \sum_{s\in S\setminus T} a_s. \tag{11}$$

(That is, any addends which are zero can be removed from a sum without changing the sum's value.) See Corollary 2.131 below for a proof of (11).

- **Renaming the index:** Let S be a finite set. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s\in S} a_s = \sum_{t\in S} a_t.$$

This is just saying that the summation index in a sum can be renamed at will, as long as its name does not clash with other notation.

- **Substituting the index I:** Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_t be an element of $\mathbb{A}$ for each $t \in T$. Then,

$$\sum_{t\in T} a_t = \sum_{s\in S} a_{f(s)}. \tag{12}$$

(The idea here is that the sum $\sum\limits_{s\in S} a_{f(s)}$ contains the same addends as the sum $\sum\limits_{t\in T} a_t$.) A rigorous proof of (12) can be found in Theorem 2.132 below.

Examples:

  - For any $n \in \mathbb{N}$, we have

$$\sum_{t\in\{1,2,\ldots,n\}} t^3 = \sum_{s\in\{-n,-n+1,\ldots,-1\}} (-s)^3 .$$

(This follows from (12), applied to $S = \{-n, -n+1, \ldots, -1\}$, $T = \{1, 2, \ldots, n\}$, $f(s) = -s$, and $a_t = t^3$.)

– The sets S and T in (12) may well be the same. For example, for any $n \in \mathbb{N}$, we have

$$\sum_{t \in \{1,2,\ldots,n\}} t^3 = \sum_{s \in \{1,2,\ldots,n\}} (n+1-s)^3 .$$

(This follows from (12), applied to $S = \{1,2,\ldots,n\}$, $T = \{1,2,\ldots,n\}$, $f(s) = n+1-s$ and $a_t = t^3$.)

– More generally: Let u and v be two integers. Then, the map $\{u, u+1, \ldots, v\} \to \{u, u+1, \ldots, v\}$ sending each $s \in \{u, u+1, \ldots, v\}$ to $u+v-s$ is a bijection[20]. Hence, we can substitute $u+v-s$ for s in the sum $\sum_{s=u}^{v} a_s$ whenever an element a_s of $\mathbb{A}$ is given for each $s \in \{u, u+1, \ldots, v\}$. We thus obtain the formula

$$\sum_{s=u}^{v} a_s = \sum_{s=u}^{v} a_{u+v-s}.$$

Remark:

– When I use (12) to rewrite the sum $\sum_{t \in T} a_t$ as $\sum_{s \in S} a_{f(s)}$, I say that I have "substituted $f(s)$ for t in the sum". Conversely, when I use (12) to rewrite the sum $\sum_{s \in S} a_{f(s)}$ as $\sum_{t \in T} a_t$, I say that I have "substituted t for $f(s)$ in the sum".

– For convenience, I have chosen s and t as summation indices in (12). But as before, they can be chosen to be any letters not otherwise used. It is perfectly okay to use one and the same letter for both of them, e.g., to write $\sum_{s \in T} a_s = \sum_{s \in S} a_{f(s)}$.

– Here is the probably most famous example of substitution in a sum: Fix a nonnegative integer n. Then, we can substitute $n-i$ for i in the sum $\sum_{i=0}^{n} i$ (since the map $\{0,1,\ldots,n\} \to \{0,1,\ldots,n\}$, $i \mapsto n-i$ is a bijection). Thus, we obtain

$$\sum_{i=0}^{n} i = \sum_{i=0}^{n} (n-i) .$$

[20] Check this!

Now,

$$\begin{aligned}
2\sum_{i=0}^{n} i &= \sum_{i=0}^{n} i + \underbrace{\sum_{i=0}^{n} i}_{=\sum_{i=0}^{n}(n-i)} && \left(\text{since } 2q = q+q \text{ for every } q \in \mathbb{Q}\right) \\
&= \sum_{i=0}^{n} i + \sum_{i=0}^{n} (n-i) \\
&= \sum_{i=0}^{n} \underbrace{(i + (n-i))}_{=n} && \left(\text{here, we have used (8) backwards}\right) \\
&= \sum_{i=0}^{n} n = (n+1)\, n && \left(\text{by (6)}\right) \\
&= n\,(n+1),
\end{aligned}$$

and therefore

$$\sum_{i=0}^{n} i = \frac{n\,(n+1)}{2}. \tag{13}$$

Since $\sum_{i=0}^{n} i = 0 + \sum_{i=1}^{n} i = \sum_{i=1}^{n} i$, this rewrites as

$$\sum_{i=1}^{n} i = \frac{n\,(n+1)}{2}. \tag{14}$$

This is the famous "Little Gauss formula" (supposedly discovered by Carl Friedrich Gauss in primary school, but already known to the Pythagoreans).

- **Substituting the index II:** Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s\in S} a_s = \sum_{t\in T} a_{f^{-1}(t)}. \tag{15}$$

This is, of course, just (12) but applied to T, S and f^{-1} instead of S, T and f. (Nevertheless, I prefer to mention (15) separately because it often is used in this very form.)

- **Telescoping sums:** Let u and v be two integers such that $u-1 \le v$. Let a_s be an element of $\mathbb{A}$ for each $s \in \{u-1, u, \ldots, v\}$. Then,

$$\sum_{s=u}^{v} (a_s - a_{s-1}) = a_v - a_{u-1}. \tag{16}$$

Examples:

– Let us give a new proof of (14). Indeed, fix a nonnegative integer n. An easy computation reveals that

$$s = \frac{s(s+1)}{2} - \frac{(s-1)((s-1)+1)}{2} \tag{17}$$

for each $s \in \mathbb{Z}$. Thus,

$$\begin{aligned}
\sum_{i=1}^{n} i &= \sum_{s=1}^{n} s = \sum_{s=1}^{n} \left(\frac{s(s+1)}{2} - \frac{(s-1)((s-1)+1)}{2} \right) \qquad \text{(by (17))} \\
&= \frac{n(n+1)}{2} - \underbrace{\frac{(1-1)((1-1)+1)}{2}}_{=0} \\
&\qquad \left(\text{by (16), applied to } u = 1,\ v = n \text{ and } a_s = \frac{s(s+1)}{2} \right) \\
&= \frac{n(n+1)}{2}.
\end{aligned}$$

Thus, (14) is proven again. This kind of proof works often when we need to prove a formula like (14); the only tricky part was to "guess" the right value of a_s, which is straightforward if you know what you are looking for (you want $a_n - a_0$ to be $\frac{n(n+1)}{2}$), but rather tricky if you don't.

– Another application of (16) is a proof of the well-known formula

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} \qquad \text{for all } n \in \mathbb{N}.$$

Indeed, an easy computation reveals that

$$s^2 = \frac{s(s+1)(2s+1)}{6} - \frac{(s-1)((s-1)+1)(2(s-1)+1)}{6}$$

for each $s \in \mathbb{Z}$; now, as in the previous example, we can sum this equality over all $s \in \{1, 2, \ldots, n\}$ and apply (16) to obtain our claim.

– Here is another important identity that follows from (16): If a and b are any elements of $\mathbb{A}$, and if $m \in \mathbb{N}$, then

$$(a-b) \sum_{i=0}^{m-1} a^i b^{m-1-i} = a^m - b^m. \tag{18}$$

(This is one of the versions of the "geometric series formula".) To prove

(18), we observe that

$$\begin{aligned}
&(a-b)\sum_{i=0}^{m-1} a^i b^{m-1-i} \\
&= \sum_{i=0}^{m-1} \underbrace{(a-b)\,a^i b^{m-1-i}}_{=aa^ib^{m-1-i}-ba^ib^{m-1-i}} \qquad \text{(this follows from (9))} \\
&= \sum_{i=0}^{m-1} \left(aa^i b^{m-1-i} - \underbrace{ba^i}_{=a^ib} b^{m-1-i} \right) \\
&= \sum_{i=0}^{m-1} \left(\underbrace{aa^i}_{=a^{i+1}} b^{m-1-i} - \underbrace{a^i}_{\substack{=a^{(i-1)+1} \\ (\text{since } i=(i-1)+1)}} \underbrace{bb^{m-1-i}}_{\substack{=b^{(m-1-i)+1}=b^{m-1-(i-1)} \\ (\text{since } (m-1-i)+1=m-1-(i-1))}} \right) \\
&= \sum_{i=0}^{m-1} \left(a^{i+1} b^{m-1-i} - a^{(i-1)+1} b^{m-1-(i-1)} \right) \\
&= \sum_{s=0}^{m-1} \left(a^{s+1} b^{m-1-s} - a^{(s-1)+1} b^{m-1-(s-1)} \right) \\
&\qquad \text{(here, we have renamed the summation index } i \text{ as } s) \\
&= \underbrace{a^{(m-1)+1}}_{\substack{=a^m \\ (\text{since } (m-1)+1=m)}} \underbrace{b^{m-1-(m-1)}}_{\substack{=b^0 \\ (\text{since } m-1-(m-1)=0)}} - \underbrace{a^{(0-1)+1}}_{\substack{=a^0 \\ (\text{since } (0-1)+1=0)}} \underbrace{b^{m-1-(0-1)}}_{\substack{=b^m \\ (\text{since } m-1-(0-1)=m)}} \\
&\qquad \left(\text{by (16) (applied to } u=0,\ v=m-1 \text{ and } a_s = a^{s+1}b^{m-1-s}) \right) \\
&= a^m \underbrace{b^0}_{=1} - \underbrace{a^0}_{=1} b^m = a^m - b^m.
\end{aligned}$$

– Other examples for the use of (16) can be found on the Wikipedia page for "telescoping series". Let me add just one more example: Given $n \in \mathbb{N}$, we want to compute $\sum_{i=1}^{n} \dfrac{1}{\sqrt{i}+\sqrt{i+1}}$. (Here, of course, we need to take $\mathbb{A} = \mathbb{R}$ or $\mathbb{A} = \mathbb{C}$.) We proceed as follows: For every positive integer i, we have

$$\frac{1}{\sqrt{i}+\sqrt{i+1}} = \frac{\left(\sqrt{i+1}-\sqrt{i}\right)}{\left(\sqrt{i}+\sqrt{i+1}\right)\left(\sqrt{i+1}-\sqrt{i}\right)} = \sqrt{i+1}-\sqrt{i}$$

(since $\left(\sqrt{i}+\sqrt{i+1}\right)\left(\sqrt{i+1}-\sqrt{i}\right) = \left(\sqrt{i+1}\right)^2 - \left(\sqrt{i}\right)^2 = (i+1) -$

$i = 1$). Thus,

$$\sum_{i=1}^{n} \frac{1}{\sqrt{i} + \sqrt{i+1}}$$
$$= \sum_{i=1}^{n} \left(\sqrt{i+1} - \sqrt{i}\right) = \sum_{s=2}^{n+1} \left(\sqrt{s} - \sqrt{s-1}\right)$$
$$\left(\begin{array}{c}\text{here, we have substituted } s-1 \text{ for } i \text{ in the sum,}\\ \text{since the map } \{2,3,\ldots,n+1\} \to \{1,2,\ldots,n\},\ s \mapsto s-1\\ \text{is a bijection}\end{array}\right)$$
$$= \sqrt{n+1} - \underbrace{\sqrt{2-1}}_{=\sqrt{1}=1}$$
$$\left(\text{by (16), applied to } u = 2,\ v = n+1 \text{ and } a_s = \sqrt{s} - \sqrt{s-1}\right)$$
$$= \sqrt{n+1} - 1.$$

Remarks:

– When we use the equality (16) to rewrite the sum $\sum_{s=u}^{v} (a_s - a_{s-1})$ as $a_v - a_{u-1}$, we can say that the sum $\sum_{s=u}^{v} (a_s - a_{s-1})$ "telescopes" to $a_v - a_{u-1}$. A sum like $\sum_{s=u}^{v} (a_s - a_{s-1})$ is said to be a "telescoping sum". This terminology references the idea that the sum $\sum_{s=u}^{v} (a_s - a_{s-1})$ "shrink" to the simple difference $a_v - a_{u-1}$ like a telescope does when it is collapsed.

– Here is a *proof of (16):* Let u and v be two integers such that $u - 1 \leq v$. Let a_s be an element of $\mathbb{A}$ for each $s \in \{u-1, u, \ldots, v\}$. Then, (8) (applied to $a_s - a_{s-1}$ and a_{s-1} instead of a_s and b_s) yields

$$\sum_{s=u}^{v} ((a_s - a_{s-1}) + a_{s-1}) = \sum_{s=u}^{v} (a_s - a_{s-1}) + \sum_{s=u}^{v} a_{s-1}.$$

Solving this equation for $\sum_{s=u}^{v} (a_s - a_{s-1})$, we obtain

$$\sum_{s=u}^{v} (a_s - a_{s-1}) = \sum_{s=u}^{v} \underbrace{((a_s - a_{s-1}) + a_{s-1})}_{=a_s} - \underbrace{\sum_{s=u}^{v} a_{s-1}}_{\substack{= \sum_{s=u-1}^{v-1} a_s \\ \text{(here, we have substituted } s \text{ for } s-1 \\ \text{in the sum)}}}$$
$$= \sum_{s=u}^{v} a_s - \sum_{s=u-1}^{v-1} a_s. \tag{19}$$

But $u-1 \leq v$. Hence, we can split off the addend for $s=u-1$ from the sum $\sum_{s=u-1}^{v} a_s$. We thus obtain

$$\sum_{s=u-1}^{v} a_s = a_{u-1} + \sum_{s=u}^{v} a_s.$$

Solving this equation for $\sum_{s=u}^{v} a_s$, we obtain

$$\sum_{s=u}^{v} a_s = \sum_{s=u-1}^{v} a_s - a_{u-1}. \tag{20}$$

Also, $u-1 \leq v$. Hence, we can split off the addend for $s=v$ from the sum $\sum_{s=u-1}^{v} a_s$. We thus obtain

$$\sum_{s=u-1}^{v} a_s = a_v + \sum_{s=u-1}^{v-1} a_s.$$

Solving this equation for $\sum_{s=u-1}^{v-1} a_s$, we obtain

$$\sum_{s=u-1}^{v-1} a_s = \sum_{s=u-1}^{v} a_s - a_v. \tag{21}$$

Now, (19) becomes

$$\begin{aligned}\sum_{s=u}^{v} (a_s - a_{s-1}) &= \underbrace{\sum_{s=u}^{v} a_s}_{\substack{=\sum_{s=u-1}^{v} a_s - a_{u-1} \\ \text{(by (20))}}} - \underbrace{\sum_{s=u-1}^{v-1} a_s}_{\substack{=\sum_{s=u-1}^{v} a_s - a_v \\ \text{(by (21))}}} \\ &= \left(\sum_{s=u-1}^{v} a_s - a_{u-1}\right) - \left(\sum_{s=u-1}^{v} a_s - a_v\right) = a_v - a_{u-1}.\end{aligned}$$

This proves (16).

- **Restricting to a subset:** Let S be a finite set. Let T be a subset of S. Let a_s be an element of $\mathbb{A}$ for each $s \in T$. Then,

$$\sum_{\substack{s\in S;\\ s\in T}} a_s = \sum_{s\in T} a_s.$$

This is because the $s \in S$ satisfying $s \in T$ are exactly the elements of T.

Remark: Here is a slightly more general form of this rule: Let S be a finite set. Let T be a subset of S. Let $\mathcal{A}(s)$ be a logical statement for each $s \in S$. Let a_s be an element of $\mathbb{A}$ for each $s \in T$ satisfying $\mathcal{A}(s)$. Then,

$$\sum_{\substack{s \in S; \\ s \in T; \\ \mathcal{A}(s)}} a_s = \sum_{\substack{s \in T; \\ \mathcal{A}(s)}} a_s.$$

- **Splitting a sum by a value of a function:** Let S be a finite set. Let W be a set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s \in S} a_s = \sum_{w \in W} \sum_{\substack{s \in S; \\ f(s) = w}} a_s. \qquad (22)$$

The idea behind this formula is the following: The left hand side is the sum of all a_s for $s \in S$. The right hand side is the same sum, but split in a particular way: First, for each $w \in W$, we sum the a_s for all $s \in S$ satisfying $f(s) = w$, and then we take the sum of all these "partial sums". For a rigorous proof of (22), see Theorem 2.127 (for the case when W is finite) and Theorem 2.147 (for the general case).

Examples:

– Let $n \in \mathbb{N}$. Then,

$$\sum_{s \in \{-n, -(n-1), \ldots, n\}} s^3 = \sum_{w \in \{0,1,\ldots,n\}} \sum_{\substack{s \in \{-n, -(n-1), \ldots, n\}; \\ |s| = w}} s^3. \qquad (23)$$

(This follows from (22), applied to $S = \{-n, -(n-1), \ldots, n\}$, $W = \{0, 1, \ldots, n\}$ and $f(s) = |s|$.) You might wonder what you gain by this observation. But actually, it allows you to compute the sum: For any $w \in \{0, 1, \ldots, n\}$, the sum $\sum_{\substack{s \in \{-n, -(n-1), \ldots, n\}; \\ |s| = w}} s^3$ is 0 [21], and therefore (23) becomes

$$\sum_{s \in \{-n, -(n-1), \ldots, n\}} s^3 = \sum_{w \in \{0,1,\ldots,n\}} \underbrace{\sum_{\substack{s \in \{-n, -(n-1), \ldots, n\}; \\ |s| = w}} s^3}_{=0} = \sum_{w \in \{0,1,\ldots,n\}} 0 = 0.$$

Thus, a strategic application of (22) can help in evaluating a sum.

[21] *Proof.* If $w = 0$, then this sum $\sum_{\substack{s \in \{-n, -(n-1), \ldots, n\}; \\ |s| = w}} s^3$ consists of one addend only, and this addend is 0^3. If $w > 0$, then this sum has two addends, namely $(-w)^3$ and w^3. In either case, the sum is 0 (because $0^3 = 0$ and $(-w)^3 + w^3 = -w^3 + w^3 = 0$).

– Let S be a finite set. Let W be a set. Let $f : S \to W$ be a map. If we apply (22) to $a_s = 1$, then we obtain

$$\sum_{s\in S} 1 = \sum_{w\in W} \underbrace{\sum_{\substack{s\in S;\\ f(s)=w}} 1}_{\substack{=\left|\{s\in S \ | \ f(s)=w\}\right|\cdot 1\\ =\left|\{s\in S \ | \ f(s)=w\}\right|}} = \sum_{w\in W} \left|\{s \in S \ | \ f(s) = w\}\right|.$$

Since $\sum_{s\in S} 1 = |S| \cdot 1 = |S|$, this rewrites as follows:

$$|S| = \sum_{w\in W} \left|\{s \in S \ | \ f(s) = w\}\right|. \qquad (24)$$

This equality is often called the *shepherd's principle*, because it is connected to the joke that "in order to count a flock of sheep, just count the legs and divide by 4". The connection is somewhat weak, actually; the equality (24) is better regarded as a formalization of the (less funny) idea that in order to count all legs of a flock of sheep, you can count the legs of every single sheep, and then sum the resulting numbers over all sheep in the flock. Think of the S in (24) as the set of all legs of all sheep in the flock; think of W as the set of all sheep in the flock; and think of f as the function which sends every leg to the (hopefully uniquely determined) sheep it belongs to.

Remarks:

– If $f : S \to W$ is a map between two sets S and W, and if w is an element of W, then it is common to denote the set $\{s \in S \ | \ f(s) = w\}$ by $f^{-1}(w)$. (Formally speaking, this notation might clash with the notation $f^{-1}(w)$ for the actual preimage of w when f happens to be bijective; but in practice, this causes far less confusion than it might seem to.) Using this notation, we can rewrite (22) as follows:

$$\sum_{s\in S} a_s = \sum_{w\in W} \underbrace{\sum_{\substack{s\in S;\\ f(s)=w}}}_{=\sum_{s\in f^{-1}(w)}} a_s = \sum_{w\in W} \sum_{s\in f^{-1}(w)} a_s. \qquad (25)$$

– When I rewrite a sum $\sum_{s\in S} a_s$ as $\sum_{w\in W} \sum_{\substack{s\in S;\\ f(s)=w}} a_s$ (or as $\sum_{w\in W} \sum_{s\in f^{-1}(w)} a_s$), I say that I am "splitting the sum according to the value of $f(s)$". (Though, most of the time, I shall be doing such manipulations without explicit mention.)

- **Splitting a sum into subsums:** Let S be a finite set. Let $S_1, S_2, \ldots, S_n$ be finitely many subsets of S. Assume that these subsets $S_1, S_2, \ldots, S_n$ are pairwise disjoint (i.e., we have $S_i \cap S_j = \varnothing$ for any two distinct elements i and j of $\{1, 2, \ldots, n\}$) and their union is S. (Thus, every element of S lies in precisely one of the subsets $S_1, S_2, \ldots, S_n$.) Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s \in S} a_s = \sum_{w=1}^{n} \sum_{s \in S_w} a_s. \tag{26}$$

This is a generalization of (3) (indeed, (3) is obtained from (26) by setting $n = 2$, $S_1 = X$ and $S_2 = Y$). It is also a consequence of (22): Indeed, set $W = \{1, 2, \ldots, n\}$, and define a map $f : S \to W$ to send each $s \in S$ to the unique $w \in \{1, 2, \ldots, n\}$ for which $s \in S_w$. Then, every $w \in W$ satisfies $\sum_{\substack{s \in S; \\ f(s) = w}} a_s = \sum_{s \in S_w} a_s$; therefore, (22) becomes (26).

Example: If we set $a_s = 1$ for each $s \in S$, then (26) becomes

$$\sum_{s \in S} 1 = \sum_{w=1}^{n} \underbrace{\sum_{s \in S_w} 1}_{= |S_w|} = \sum_{w=1}^{n} |S_w| .$$

Hence,

$$\sum_{w=1}^{n} |S_w| = \sum_{s \in S} 1 = |S| \cdot 1 = |S| .$$

- **Fubini's theorem (interchanging the order of summation):** Let X and Y be two finite sets. Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x, y) \in X \times Y$. Then,

$$\sum_{x \in X} \sum_{y \in Y} a_{(x,y)} = \sum_{(x,y) \in X \times Y} a_{(x,y)} = \sum_{y \in Y} \sum_{x \in X} a_{(x,y)}. \tag{27}$$

This is called *Fubini's theorem for finite sums*, and is a lot easier to prove than what analysts tend to call Fubini's theorem. I shall sketch a proof shortly (in the Remark below); but first, let me give some intuition for the statement. Imagine that you have a rectangular table filled with numbers. If you want to sum the numbers in the table, you can proceed in several ways. One way is to sum the numbers in each row, and then sum all the sums you have obtained. Another way is to sum the numbers in each column, and then sum all the obtained sums. Either way, you get the same result – namely, the sum of all numbers in the table. This is essentially what (27) says, at least when $X = \{1, 2, \ldots, n\}$ and $Y = \{1, 2, \ldots, m\}$ for some integers n and m. In this case, the numbers $a_{(x,y)}$ can be viewed as forming a table, where $a_{(x,y)}$ is placed in the cell at the intersection of row x with column y. When X and Y are arbitrary finite sets (not necessarily $\{1, 2, \ldots, n\}$ and $\{1, 2, \ldots, m\}$), then you need to slightly stretch your imagination in order to see the $a_{(x,y)}$ as "forming a table"; in fact, there is no obvious order in which the numbers appear in a row or column, but there is still a notion of rows and columns.

Examples:

– Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x,y) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,m\}$. Then,

$$\sum_{x=1}^{n} \sum_{y=1}^{m} a_{(x,y)} = \sum_{(x,y) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,m\}} a_{(x,y)} = \sum_{y=1}^{m} \sum_{x=1}^{n} a_{(x,y)}. \qquad (28)$$

(This follows from (27), applied to $X = \{1,2,\ldots,n\}$ and $Y = \{1,2,\ldots,m\}$.) We can rewrite the equality (28) without using $\sum$ signs; it then takes the following form:

$$\begin{aligned}
&\left(a_{(1,1)} + a_{(1,2)} + \cdots + a_{(1,m)}\right) \\
&\qquad + \left(a_{(2,1)} + a_{(2,2)} + \cdots + a_{(2,m)}\right) \\
&\qquad + \cdots \\
&\qquad + \left(a_{(n,1)} + a_{(n,2)} + \cdots + a_{(n,m)}\right) \\
&= a_{(1,1)} + a_{(1,2)} + \cdots + a_{(n,m)} \qquad \left(\text{this is the sum of all } nm \text{ numbers } a_{(x,y)}\right) \\
&= \left(a_{(1,1)} + a_{(2,1)} + \cdots + a_{(n,1)}\right) \\
&\qquad + \left(a_{(1,2)} + a_{(2,2)} + \cdots + a_{(n,2)}\right) \\
&\qquad + \cdots \\
&\qquad + \left(a_{(1,m)} + a_{(2,m)} + \cdots + a_{(n,m)}\right).
\end{aligned}$$

In other words, we can sum the entries of the rectangular table

$$\begin{array}{|cccc|}
\hline
a_{(1,1)} & a_{(1,2)} & \cdots & a_{(1,m)} \\
a_{(2,1)} & a_{(2,2)} & \cdots & a_{(2,m)} \\
\vdots & \vdots & \ddots & \vdots \\
a_{(n,1)} & a_{(n,2)} & \cdots & a_{(n,m)} \\
\hline
\end{array}$$

in three different ways:

(a) row by row (i.e., first summing the entries in each row, then summing up the n resulting tallies);

(b) arbitrarily (i.e., just summing all entries of the table in some arbitrary order);

(c) column by column (i.e., first summing the entries in each column, then summing up the m resulting tallies);

and each time, we get the same result.

– Here is a concrete application of (28): Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. We want to compute $\sum_{(x,y)\in\{1,2,\ldots,n\}\times\{1,2,\ldots,m\}} xy$. (This is the sum of all entries of the $n \times m$ multiplication table.) Applying (28) to $a_{(x,y)} = xy$, we obtain

$$\sum_{x=1}^{n}\sum_{y=1}^{m} xy = \sum_{(x,y)\in\{1,2,\ldots,n\}\times\{1,2,\ldots,m\}} xy = \sum_{y=1}^{m}\sum_{x=1}^{n} xy.$$

Hence,

$$\begin{aligned}
&\sum_{(x,y)\in\{1,2,\ldots,n\}\times\{1,2,\ldots,m\}} xy \\
&= \sum_{x=1}^{n} \underbrace{\sum_{y=1}^{m} xy}_{\substack{=\sum_{s=1}^{m} xs = x\sum_{s=1}^{m} s \\ \text{(by (9), applied to } S=\{1,2,\ldots,m\}, \\ a_s=s \text{ and } \lambda=x)}} = \sum_{x=1}^{n} x \underbrace{\sum_{s=1}^{m} s}_{\substack{=\sum_{i=1}^{m} i = \frac{m(m+1)}{2} \\ \text{(by (14), applied to } m \\ \text{instead of } n)}} \\
&= \sum_{x=1}^{n} x\frac{m(m+1)}{2} = \sum_{x=1}^{n} \frac{m(m+1)}{2}x = \sum_{s=1}^{n} \frac{m(m+1)}{2}s \\
&= \frac{m(m+1)}{2} \underbrace{\sum_{s=1}^{n} s}_{\substack{=\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \\ \text{(by (14))}}} \\
&\qquad \left(\text{by (9), applied to } S = \{1,2,\ldots,n\}, a_s = s \text{ and } \lambda = \frac{m(m+1)}{2}\right) \\
&= \frac{m(m+1)}{2} \cdot \frac{n(n+1)}{2}.
\end{aligned}$$

Remarks:

– I have promised to outline a proof of (27). Here it comes: Let $S = X \times Y$ and $W = Y$, and let $f : S \to W$ be the map which sends every pair (x, y) to its second entry y. Then, (25) shows that

$$\sum_{s\in X\times Y} a_s = \sum_{w\in Y}\sum_{s\in f^{-1}(w)} a_s. \tag{29}$$

But for every given $w \in Y$, the set $f^{-1}(w)$ is simply the set of all pairs (x, w) with $x \in X$. Thus, for every given $w \in Y$, there is a bijection $g_w : X \to f^{-1}(w)$ given by

$$g_w(x) = (x, w) \qquad \text{for all } x \in X.$$

Hence, for every given $w \in Y$, we can substitute $g_w(x)$ for s in the sum $\sum_{s \in f^{-1}(w)} a_s$, and thus obtain

$$\sum_{s \in f^{-1}(w)} a_s = \sum_{x \in X} \underbrace{a_{g_w(x)}}_{\substack{= a_{(x,w)} \\ (\text{since } g_w(x) = (x,w))}} = \sum_{x \in X} a_{(x,w)}.$$

Hence, (29) becomes

$$\sum_{s \in X \times Y} a_s = \sum_{w \in Y} \underbrace{\sum_{s \in f^{-1}(w)} a_s}_{= \sum_{x \in X} a_{(x,w)}} = \sum_{w \in Y} \sum_{x \in X} a_{(x,w)} = \sum_{y \in Y} \sum_{x \in X} a_{(x,y)}$$

(here, we have renamed the summation index w as y in the outer sum). Therefore,

$$\sum_{y \in Y} \sum_{x \in X} a_{(x,y)} = \sum_{s \in X \times Y} a_s = \sum_{(x,y) \in X \times Y} a_{(x,y)}$$

(here, we have renamed the summation index s as (x, y)). Thus, we have proven the second part of the equality (27). The first part can be proven similarly.

– I like to abbreviate the equality (28) as follows:

$$\sum_{x=1}^{n} \sum_{y=1}^{m} = \sum_{(x,y) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,m\}} = \sum_{y=1}^{m} \sum_{x=1}^{n}. \tag{30}$$

This is an "equality between summation signs"; it should be understood as follows: Every time you see an "$\sum_{x=1}^{n} \sum_{y=1}^{m}$" in an expression, you can replace it by a "$\sum_{(x,y) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,m\}}$" or by a "$\sum_{y=1}^{m} \sum_{x=1}^{n}$", and similarly the other ways round.

- **Triangular Fubini's theorem I:** The equality (28) formalizes the idea that we can sum the entries of a rectangular table by first tallying each row and then adding together, or first tallying each column and adding together. The same holds for triangular tables. More precisely: Let $n \in \mathbb{N}$. Let T_n be the set

$\left\{(x,y) \in \{1,2,3,\ldots\}^2 \mid x+y \leq n\right\}$. (For instance, if $n = 3$, then $T_n = T_3 = \{(1,1), (1,2), (2,1)\}$.) Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x,y) \in T_n$. Then,

$$\sum_{x=1}^{n} \sum_{y=1}^{n-x} a_{(x,y)} = \sum_{(x,y) \in T_n} a_{(x,y)} = \sum_{y=1}^{n} \sum_{x=1}^{n-y} a_{(x,y)}. \tag{31}$$

Examples:

– In the case when $n = 4$, the formula (31) (rewritten without the use of $\sum$ signs) looks as follows:

$$\begin{aligned} & \left(a_{(1,1)} + a_{(1,2)} + a_{(1,3)}\right) + \left(a_{(2,1)} + a_{(2,2)}\right) + a_{(3,1)} + (\text{empty sum}) \\ & = \left(\text{the sum of the } a_{(x,y)} \text{ for all } (x,y) \in T_4\right) \\ & = \left(a_{(1,1)} + a_{(2,1)} + a_{(3,1)}\right) + \left(a_{(1,2)} + a_{(2,2)}\right) + a_{(1,3)} + (\text{empty sum}). \end{aligned}$$

In other words, we can sum the entries of the triangular table

$a_{(1,1)}$	$a_{(1,2)}$	$a_{(1,3)}$
$a_{(2,1)}$	$a_{(2,2)}$	
$a_{(3,1)}$		

in three different ways:

(a) row by row (i.e., first summing the entries in each row, then summing up the resulting tallies);

(b) arbitrarily (i.e., just summing all entries of the table in some arbitrary order);

(c) column by column (i.e., first summing the entries in each column, then summing up the resulting tallies);

and each time, we get the same result.

– Let us use (31) to compute $|T_n|$. Indeed, we can apply (31) to $a_{(x,y)} = 1$. Thus, we obtain

$$\sum_{x=1}^{n} \sum_{y=1}^{n-x} 1 = \sum_{(x,y) \in T_n} 1 = \sum_{y=1}^{n} \sum_{x=1}^{n-y} 1.$$

Hence,

$$\sum_{x=1}^{n} \sum_{y=1}^{n-x} 1 = \sum_{(x,y) \in T_n} 1 = |T_n|,$$

so that

$$|T_n| = \sum_{x=1}^{n} \underbrace{\sum_{y=1}^{n-x} 1}_{=n-x} = \sum_{x=1}^{n} (n-x) = \sum_{i=0}^{n-1} i$$

$$\left(\begin{array}{c}\text{here, we have substituted } i \text{ for } n-x \text{ in the sum,}\\ \text{since the map } \{1,2,\ldots,n\} \to \{0,1,\ldots,n-1\},\ x \mapsto n-x\\ \text{is a bijection}\end{array}\right)$$

$$= \frac{(n-1)((n-1)+1)}{2} \qquad \text{(by (13), applied to } n-1 \text{ instead of } n\text{)}$$

$$= \frac{(n-1)n}{2}.$$

Remarks:

– The sum $\sum_{(x,y)\in T_n} a_{(x,y)}$ in (31) can also be rewritten as $\sum_{\substack{(x,y)\in\{1,2,3,\ldots\}^2;\\ x+y\le n}} a_{(x,y)}$.

– Let us prove (31). Indeed, the proof will be very similar to our proof of (27) above. Let $S = T_n$ and $W = \{1,2,\ldots,n\}$, and let $f : S \to W$ be the map which sends every pair (x,y) to its second entry y. Then, (25) shows that

$$\sum_{s\in T_n} a_s = \sum_{w\in W} \sum_{s\in f^{-1}(w)} a_s. \tag{32}$$

But for every given $w \in W$, the set $f^{-1}(w)$ is simply the set of all pairs (x,w) with $x \in \{1,2,\ldots,n-w\}$. Thus, for every given $w \in W$, there is a bijection $g_w : \{1,2,\ldots,n-w\} \to f^{-1}(w)$ given by

$$g_w(x) = (x,w) \qquad \text{for all } x \in \{1,2,\ldots,n-w\}.$$

Hence, for every given $w \in W$, we can substitute $g_w(x)$ for s in the sum $\sum_{s\in f^{-1}(w)} a_s$, and thus obtain

$$\sum_{s\in f^{-1}(w)} a_s = \underbrace{\sum_{x\in\{1,2,\ldots,n-w\}}}_{=\sum_{x=1}^{n-w}} \underbrace{a_{g_w(x)}}_{\substack{=a_{(x,w)}\\ \text{(since } g_w(x)=(x,w))}} = \sum_{x=1}^{n-w} a_{(x,w)}.$$

Hence, (32) becomes

$$\sum_{s\in T_n} a_s = \underbrace{\sum_{w\in W}}_{\substack{=\sum_{w=1}^{n}\\ \text{(since } W=\{1,2,\ldots,n\})}} \underbrace{\sum_{s\in f^{-1}(w)} a_s}_{=\sum_{x=1}^{n-w} a_{(x,w)}} = \sum_{w=1}^{n} \sum_{x=1}^{n-w} a_{(x,w)} = \sum_{y=1}^{n} \sum_{x=1}^{n-y} a_{(x,y)}$$

(here, we have renamed the summation index w as y in the outer sum). Therefore,

$$\sum_{y=1}^{n} \sum_{x=1}^{n-y} a_{(x,y)} = \sum_{s \in T_n} a_s = \sum_{(x,y) \in T_n} a_{(x,y)}.$$

Thus, we have proven the second part of the equality (31). The first part can be proven similarly.

- **Triangular Fubini's theorem II:** Here is another equality similar to (31). Let $n \in \mathbb{N}$. Let Q_n be the set $\left\{ (x,y) \in \{1,2,\ldots,n\}^2 \mid x \leq y \right\}$. (For instance, if $n = 3$, then $Q_n = Q_3 = \{(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)\}$.) Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x,y) \in Q_n$. Then,

$$\sum_{x=1}^{n} \sum_{y=x}^{n} a_{(x,y)} = \sum_{(x,y) \in Q_n} a_{(x,y)} = \sum_{y=1}^{n} \sum_{x=1}^{y} a_{(x,y)}. \tag{33}$$

Examples:

– In the case when $n = 4$, the formula (33) (rewritten without the use of $\sum$ signs) looks as follows:

$$\begin{aligned} & \left(a_{(1,1)} + a_{(1,2)} + a_{(1,3)} + a_{(1,4)} \right) \\ & \qquad + \left(a_{(2,2)} + a_{(2,3)} + a_{(2,4)} \right) \\ & \qquad + \left(a_{(3,3)} + a_{(3,4)} \right) \\ & \qquad + a_{(4,4)} \\ & = \left(\text{the sum of the } a_{(x,y)} \text{ for all } (x,y) \in Q_4 \right) \\ & = a_{(1,1)} \\ & \qquad + \left(a_{(1,2)} + a_{(2,2)} \right) \\ & \qquad + \left(a_{(1,3)} + a_{(2,3)} + a_{(3,3)} \right) \\ & \qquad + \left(a_{(1,4)} + a_{(2,4)} + a_{(3,4)} + a_{(4,4)} \right). \end{aligned}$$

In other words, we can sum the entries of the triangular table

$a_{(1,1)}$	$a_{(1,2)}$	$a_{(1,3)}$	$a_{(1,4)}$
	$a_{(2,2)}$	$a_{(2,3)}$	$a_{(2,4)}$
		$a_{(3,3)}$	$a_{(3,4)}$
			$a_{(4,4)}$

in three different ways:

(a) row by row (i.e., first summing the entries in each row, then summing up the resulting tallies);

(b) arbitrarily (i.e., just summing all entries of the table in some arbitrary order);

(c) column by column (i.e., first summing the entries in each column, then summing up the resulting tallies);

and each time, we get the same result.

– Let us use (33) to compute $|Q_n|$. Indeed, we can apply (33) to $a_{(x,y)} = 1$. Thus, we obtain

$$\sum_{x=1}^{n} \sum_{y=x}^{n} 1 = \sum_{(x,y) \in Q_n} 1 = \sum_{y=1}^{n} \sum_{x=1}^{y} 1.$$

Hence,

$$\sum_{y=1}^{n} \sum_{x=1}^{y} 1 = \sum_{(x,y) \in Q_n} 1 = |Q_n|,$$

so that

$$|Q_n| = \sum_{y=1}^{n} \underbrace{\sum_{x=1}^{y} 1}_{=y} = \sum_{y=1}^{n} y = \sum_{i=1}^{n} i = \frac{n(n+1)}{2} \qquad \text{(by (14))}.$$

Remarks:

– The sum $\sum_{(x,y) \in Q_n} a_{(x,y)}$ in (33) can also be rewritten as $\sum_{\substack{(x,y) \in \{1,2,\ldots,n\}^2; \\ x \leq y}} a_{(x,y)}$.

It is also often written as $\sum_{1 \leq x \leq y \leq n} a_{(x,y)}$.

– The proof of (33) is similar to that of (31).

• **Fubini's theorem with a predicate:** Let X and Y be two finite sets. For every pair $(x, y) \in X \times Y$, let $\mathcal{A}(x, y)$ be a logical statement. For each $(x, y) \in X \times Y$ satisfying $\mathcal{A}(x, y)$, let $a_{(x,y)}$ be an element of $\mathbb{A}$. Then,

$$\sum_{x \in X} \sum_{\substack{y \in Y; \\ \mathcal{A}(x,y)}} a_{(x,y)} = \sum_{\substack{(x,y) \in X \times Y; \\ \mathcal{A}(x,y)}} a_{(x,y)} = \sum_{y \in Y} \sum_{\substack{x \in X; \\ \mathcal{A}(x,y)}} a_{(x,y)}. \qquad (34)$$

Examples:

– For any $n \in \mathbb{N}$ and $m \in \mathbb{N}$, we have

$$\sum_{x \in \{1,2,\ldots,n\}} \sum_{\substack{y \in \{1,2,\ldots,m\}; \\ x+y \text{ is even}}} xy = \sum_{\substack{(x,y) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,m\}; \\ x+y \text{ is even}}} xy$$
$$= \sum_{y \in \{1,2,\ldots,m\}} \sum_{\substack{x \in \{1,2,\ldots,n\}; \\ x+y \text{ is even}}} xy.$$

(This follows from (34), applied to $X = \{1,2,\ldots,n\}$, $Y = \{1,2,\ldots,m\}$ and $\mathcal{A}(x,y) = (\text{"}x+y \text{ is even"})$.)

Remarks:

– We have assumed that the sets X and Y are finite. But (34) is still valid if we replace this assumption by the weaker assumption that only finitely many $(x,y) \in X \times Y$ satisfy $\mathcal{A}(x,y)$.

– It is not hard to prove (34) by suitably adapting our proof of (27).

– The equality (31) can be derived from (34) by setting $X = \{1,2,\ldots,n\}$, $Y = \{1,2,\ldots,n\}$ and $\mathcal{A}(x,y) = (\text{"}x+y \leq n\text{"})$. Similarly, the equality (33) can be derived from (34) by setting $X = \{1,2,\ldots,n\}$, $Y = \{1,2,\ldots,n\}$ and $\mathcal{A}(x,y) = (\text{"}x \leq y\text{"})$.

• **Interchange of predicates:** Let S be a finite set. For every $s \in S$, let $\mathcal{A}(s)$ and $\mathcal{B}(s)$ be two equivalent logical statements. ("Equivalent" means that $\mathcal{A}(s)$ holds if and only if $\mathcal{B}(s)$ holds.) Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s = \sum_{\substack{s \in S; \\ \mathcal{B}(s)}} a_s.$$

(If you regard equivalent logical statements as identical, then you will see this as a tautology. If not, it is still completely obvious, since the equivalence of $\mathcal{A}(s)$ with $\mathcal{B}(s)$ shows that $\{t \in S \mid \mathcal{A}(t)\} = \{t \in S \mid \mathcal{B}(t)\}$.)

• **Substituting the index I with a predicate:** Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_t be an element of $\mathbb{A}$ for each $t \in T$. For every $t \in T$, let $\mathcal{A}(t)$ be a logical statement. Then,

$$\sum_{\substack{t \in T; \\ \mathcal{A}(t)}} a_t = \sum_{\substack{s \in S; \\ \mathcal{A}(f(s))}} a_{f(s)}. \tag{35}$$

Remarks:

– The equality (35) is a generalization of (12). There is a similar generalization of (15).

– The equality (35) can be easily derived from (12). Indeed, let S' be the subset $\{s \in S \mid \mathcal{A}(f(s))\}$ of S, and let T' be the subset $\{t \in T \mid \mathcal{A}(t)\}$ of T. Then, the map $S' \to T'$, $s \mapsto f(s)$ is well-defined and a bijection[22], and thus (12) (applied to S', T' and this map instead of S, T and f) yields $\sum\limits_{t \in T'} a_t = \sum\limits_{s \in S'} a_{f(s)}$. But this is precisely the equality (35), because clearly we have $\sum\limits_{t \in T'} = \sum\limits_{\substack{t \in T; \\ \mathcal{A}(t)}}$ and $\sum\limits_{s \in S'} = \sum\limits_{\substack{s \in S; \\ \mathcal{A}(f(s))}}$.

1.4.3. Definition of $\prod$

We shall now define the $\prod$ sign. Since the $\prod$ sign is (in many aspects) analogous to the $\sum$ sign, we shall be brief and confine ourselves to the bare necessities; we trust the reader to transfer most of what we said about $\sum$ to the case of $\prod$. In particular, we shall give very few examples and no proofs.

- If S is a finite set, and if a_s is an element of $\mathbb{A}$ for each $s \in S$, then $\prod_{s \in S} a_s$ denotes the product of all of these elements a_s. Formally, this product is defined by recursion on $|S|$, as follows:
  – If $|S| = 0$, then $\prod\limits_{s \in S} a_s$ is defined to be 1.
  – Let $n \in \mathbb{N}$. Assume that we have defined $\prod\limits_{s \in S} a_s$ for every finite set S with $|S| = n$ (and every choice of elements a_s of $\mathbb{A}$). Now, if S is a finite set with $|S| = n+1$ (and if $a_s \in \mathbb{A}$ are chosen for all $s \in S$), then $\prod\limits_{s \in S} a_s$ is defined by picking any $t \in S$ and setting
  $$\prod_{s \in S} a_s = a_t \cdot \prod_{s \in S \setminus \{t\}} a_s. \qquad (36)$$
  As for $\sum\limits_{s \in S} a_s$, this definition is not obviously legitimate, but it can be proven to be legitimate nevertheless. (The proof is analogous to the proof for $\sum\limits_{s \in S} a_s$; see Subsection 2.14.14 for details.)

  Examples:

  – If $S = \{1, 2, \ldots, n\}$ (for some $n \in \mathbb{N}$) and $a_s = s$ for every $s \in S$, then $\prod\limits_{s \in S} a_s = \prod\limits_{s \in S} s = 1 \cdot 2 \cdot \cdots \cdot n$. This number $1 \cdot 2 \cdot \cdots \cdot n$ is denoted by $n!$ and called the *factorial of n*.

[22] This is easy to see.

In particular,

$$\begin{aligned} 0! &= \prod_{s\in\{1,2,\ldots,0\}} s = \prod_{s\in\varnothing} s \qquad (\text{since } \{1,2,\ldots,0\} = \varnothing) \\ &= 1 \qquad (\text{since } |\varnothing| = 0); \\ 1! &= \prod_{s\in\{1,2,\ldots,1\}} s = \prod_{s\in\{1\}} s = 1; \\ 2! &= \prod_{s\in\{1,2,\ldots,2\}} s = \prod_{s\in\{1,2\}} s = 1\cdot 2 = 2; \\ 3! &= \prod_{s\in\{1,2,\ldots,3\}} s = \prod_{s\in\{1,2,3\}} s = 1\cdot 2\cdot 3 = 6; \end{aligned}$$

similarly,

$$4! = 1\cdot 2\cdot 3\cdot 4 = 24; \qquad 5! = 120; \qquad 6! = 720; \qquad 7! = 5040.$$

Notice that

$$n! = n\cdot(n-1)! \qquad \text{for any positive integer } n. \tag{37}$$

(This can be obtained from (36), applied to $S = \{1,2,\ldots,n\}$, $a_s = s$ and $t = n$.)

Remarks:

– The product $\prod_{s\in S} a_s$ is usually pronounced "product of the a_s over all $s \in S$" or "product of the a_s with s ranging over S" or "product of the a_s with s running through all elements of S". The letter "s" in the product is called the "product index", and its exact choice is immaterial, as long as it does not already have a different meaning outside of the product. The sign $\prod$ itself is called "the product sign" or "the $\prod$ sign". The numbers a_s are called the *factors* of the product $\prod_{s\in S} a_s$. More precisely, for any given $t \in S$, we can refer to the number a_t as the "factor corresponding to the index t" (or as the "factor for $s = t$", or as the "factor for t") of the product $\prod_{s\in S} a_s$.

– When the set S is empty, the product $\prod_{s\in S} a_s$ is called an *empty product*. Our definition implies that any empty product is 1. This convention is used throughout mathematics, except in rare occasions where a slightly subtler version of it is used[23].

[23] Just as with sums, the subtlety lies in the fact that mathematicians sometimes want an empty product to be not the integer 1 but the unity of some ring. As before, this does not matter for us right now.

– If $a \in \mathbb{A}$ and $n \in \mathbb{N}$, then the n-th power of a (written a^n) is defined by

$$a^n = \underbrace{a \cdot a \cdot \cdots \cdot a}_{n \text{ times}} = \prod_{i \in \{1,2,\ldots,n\}} a.$$

Thus, a^0 is an empty product, and therefore equal to 1. This holds for any $a \in \mathbb{A}$, including 0; thus, $0^0 = 1$. **There is nothing controversial about the equality** $0^0 = 1$; it is a consequence of the only reasonable definition of the n-th power of a number. Ignore anyone who tells you that 0^0 is "undefined" or "indeterminate" or "can be 0 or 1 or anything, depending on the context".[24]

– The product index (just like a summation index) needs not be a single letter; it can be a pair or a triple, for example.

– Mathematicians don't seem to have reached an agreement on the operator precedence of the $\prod$ sign. My convention is that the product sign has higher precedence than the plus sign (so an expression like $\prod_{s\in S} a_s + b$ must be read as $\left(\prod_{s\in S} a_s\right) + b$, and not as $\prod_{s\in S} (a_s + b)$); this is, of course, in line with the standard convention that multiplication-like operations have higher precedence than addition-like operations ("PEMDAS"). Be warned that some authors disagree even with this convention. I strongly advise against writing things like $\prod_{s\in S} a_s b$, since it might mean both $\left(\prod_{s\in S} a_s\right) b$ and $\prod_{s\in S} (a_s b)$ depending on the weather. In particular, I advise against writing things like $\prod_{s\in S} a_s \cdot \prod_{s\in S} b_s$ without parentheses (although I do use a similar convention for sums, namely $\sum_{s\in S} a_s + \sum_{s\in S} b_s$, and I find it to be fairly harmless). These rules are not carved in stone, and you should use whatever conventions make **you** safe from ambiguity; either way, you should keep in mind that other authors make different choices.

– An expression of the form "$\prod_{s\in S} a_s$" (where S is a finite set) is called a *finite product*.

– We have required the set S to be finite when defining $\prod_{s\in S} a_s$. Such products are not generally defined when S is infinite. However, **some** infinite products can be made sense of. The simplest case is when the set S might be infinite, but only finitely many among the a_s are distinct from 1. In this case, we can define $\prod_{s\in S} a_s$ simply by discarding the factors which

[24]I am talking about the **number** 0^0 here. There is also something called "the indeterminate form 0^0", which is a much different story.

equal 1 and multiplying the finitely many remaining factors. Other situations in which infinite products make sense appear in analysis and in topological algebra.

– The product $\prod_{s\in S} a_s$ always belongs to $\mathbb{A}$.

• A slightly more complicated version of the product sign is the following: Let S be a finite set, and let $\mathcal{A}(s)$ be a logical statement defined for every $s \in S$. For each $s \in S$ satisfying $\mathcal{A}(s)$, let a_s be an element of $\mathbb{A}$. Then, the product $\prod_{\substack{s\in S;\\ \mathcal{A}(s)}} a_s$ is defined by

$$\prod_{\substack{s\in S;\\ \mathcal{A}(s)}} a_s = \prod_{s\in\{t\in S \mid \mathcal{A}(t)\}} a_s.$$

• Finally, here is the simplest version of the product sign: Let u and v be two integers. As before, we understand the set $\{u, u+1, \ldots, v\}$ to be empty when $u > v$. Let a_s be an element of $\mathbb{A}$ for each $s \in \{u, u+1, \ldots, v\}$. Then, $\prod_{s=u}^{v} a_s$ is defined by

$$\prod_{s=u}^{v} a_s = \prod_{s\in\{u,u+1,\ldots,v\}} a_s.$$

Examples:

– We have $\prod_{s=1}^{n} s = 1 \cdot 2 \cdot \cdots \cdot n = n!$ for each $n \in \mathbb{N}$.

Remarks:

– The product $\prod_{s=u}^{v} a_s$ is usually pronounced "product of the a_s for all s from u to v (inclusive)". It is often written $a_u \cdot a_{u+1} \cdot \cdots \cdot a_v$ (or just $a_u a_{u+1} \cdots a_v$), but this latter notation has the same drawbacks as the similar notation $a_u + a_{u+1} + \cdots + a_v$ for $\sum_{s=u}^{v} a_s$.

– The product $\prod_{s=u}^{v} a_s$ is said to be *empty* whenever $u > v$. As with sums, it does not matter how much smaller v is than u; as long as v is smaller than u, the product is empty and equals 1.

Thus we have introduced the main three forms of the product sign.

1.4.4. Properties of $\prod$

Now, let me summarize the most important properties of the $\prod$ sign. These properties mirror the properties of $\sum$ discussed before; thus, I will again be brief.

- **Splitting-off:** Let S be a finite set. Let $t \in S$. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\prod_{s\in S} a_s = a_t \cdot \prod_{s\in S\setminus\{t\}} a_s.$$

- **Splitting:** Let S be a finite set. Let X and Y be two subsets of S such that $X \cap Y = \varnothing$ and $X \cup Y = S$. (Equivalently, X and Y are two subsets of S such that each element of S lies in **exactly** one of X and Y.) Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\prod_{s\in S} a_s = \left(\prod_{s\in X} a_s\right) \cdot \left(\prod_{s\in Y} a_s\right).$$

- **Splitting using a predicate:** Let S be a finite set. Let $\mathcal{A}(s)$ be a logical statement for each $s \in S$. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\prod_{s\in S} a_s = \left(\prod_{\substack{s\in S;\\ \mathcal{A}(s)}} a_s\right) \cdot \left(\prod_{\substack{s\in S;\\ \text{not } \mathcal{A}(s)}} a_s\right).$$

- **Multiplying equal values:** Let S be a finite set. Let a be an element of $\mathbb{A}$. Then,
$$\prod_{s\in S} a = a^{|S|}.$$

- **Splitting a factor:** Let S be a finite set. For every $s \in S$, let a_s and b_s be elements of $\mathbb{A}$. Then,
$$\prod_{s\in S} (a_s b_s) = \left(\prod_{s\in S} a_s\right) \cdot \left(\prod_{s\in S} b_s\right). \qquad (38)$$

  Examples:

  - Here is a frequently used particular case of (38): Let S be a finite set. For every $s \in S$, let b_s be an element of $\mathbb{A}$. Let a be an element of $\mathbb{A}$. Then, (38) (applied to $a_s = a$) yields
$$\prod_{s\in S} (a b_s) = \underbrace{\left(\prod_{s\in S} a\right)}_{=a^{|S|}} \cdot \left(\prod_{s\in S} b_s\right) = a^{|S|} \cdot \left(\prod_{s\in S} b_s\right). \qquad (39)$$

  - Here is an even further particular case: Let S be a finite set. For every $s \in S$, let b_s be an element of $\mathbb{A}$. Then,
$$\prod_{s\in S} \underbrace{(-b_s)}_{=(-1)b_s} = \prod_{s\in S} ((-1)\, b_s) = (-1)^{|S|} \cdot \left(\prod_{s\in S} b_s\right)$$
    (by (39), applied to $a = -1$).

- **Factoring out an exponent:** Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{A}$. Also, let $\lambda \in \mathbb{N}$. Then,
$$\prod_{s\in S} a_s^\lambda = \left(\prod_{s\in S} a_s\right)^\lambda .$$
- **Factoring out an integer exponent:** Let S be a finite set. For every $s \in S$, let a_s be a nonzero element of $\mathbb{A}$. Also, let $\lambda \in \mathbb{Z}$. Then,
$$\prod_{s\in S} a_s^\lambda = \left(\prod_{s\in S} a_s\right)^\lambda .$$
Remark: Applying this to $\lambda = -1$, we obtain
$$\prod_{s\in S} a_s^{-1} = \left(\prod_{s\in S} a_s\right)^{-1} .$$
In other words,
$$\prod_{s\in S} \frac{1}{a_s} = \frac{1}{\prod_{s\in S} a_s}.$$
- **Ones multiply to one:** Let S be a finite set. Then,
$$\prod_{s\in S} 1 = 1.$$
- **Dropping ones:** Let S be a finite set. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Let T be a subset of S such that every $s \in T$ satisfies $a_s = 1$. Then,
$$\prod_{s\in S} a_s = \prod_{s\in S\setminus T} a_s.$$
- **Renaming the index:** Let S be a finite set. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\prod_{s\in S} a_s = \prod_{t\in S} a_t.$$
- **Substituting the index I:** Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_t be an element of $\mathbb{A}$ for each $t \in T$. Then,
$$\prod_{t\in T} a_t = \prod_{s\in S} a_{f(s)}.$$
- **Substituting the index II:** Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\prod_{s\in S} a_s = \prod_{t\in T} a_{f^{-1}(t)}.$$

- **Telescoping products:** Let u and v be two integers such that $u-1 \leq v$. Let a_s be an element of $\mathbb{A}$ for each $s \in \{u-1, u, \ldots, v\}$. Then,

$$\prod_{s=u}^{v} \frac{a_s}{a_{s-1}} = \frac{a_v}{a_{u-1}} \tag{40}$$

(provided that $a_{s-1} \neq 0$ for all $s \in \{u, u+1, \ldots, v\}$).

Examples:

  - Let n be a positive integer. Then,

$$\prod_{s=2}^{n} \underbrace{\left(1 - \frac{1}{s}\right)}_{= \frac{s-1}{s} = \frac{1/s}{1/(s-1)}} = \prod_{s=2}^{n} \frac{1/s}{1/(s-1)} = \frac{1/n}{1/(2-1)}$$

(by (40), applied to $u = 2$, $v = n$ and $a_s = 1/s$)

$$= \frac{1}{n}.$$

- **Restricting to a subset:** Let S be a finite set. Let T be a subset of S. Let a_s be an element of $\mathbb{A}$ for each $s \in T$. Then,

$$\prod_{\substack{s \in S; \\ s \in T}} a_s = \prod_{s \in T} a_s.$$

Remark: Here is a slightly more general form of this rule: Let S be a finite set. Let T be a subset of S. Let $\mathcal{A}(s)$ be a logical statement for each $s \in S$. Let a_s be an element of $\mathbb{A}$ for each $s \in T$ satisfying $\mathcal{A}(s)$. Then,

$$\prod_{\substack{s \in S; \\ s \in T; \\ \mathcal{A}(s)}} a_s = \prod_{\substack{s \in T; \\ \mathcal{A}(s)}} a_s.$$

- **Splitting a product by a value of a function:** Let S be a finite set. Let W be a set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\prod_{s \in S} a_s = \prod_{w \in W} \prod_{\substack{s \in S; \\ f(s) = w}} a_s.$$

(The right hand side is to be read as $\prod_{w \in W} \left(\prod_{\substack{s \in S; \\ f(s) = w}} a_s \right)$.)

- **Splitting a product into subproducts:** Let S be a finite set. Let $S_1, S_2, \ldots, S_n$ be finitely many subsets of S. Assume that these subsets $S_1, S_2, \ldots, S_n$ are pairwise disjoint (i.e., we have $S_i \cap S_j = \varnothing$ for any two distinct elements i and j of $\{1, 2, \ldots, n\}$) and their union is S. (Thus, every element of S lies in precisely one of the subsets $S_1, S_2, \ldots, S_n$.) Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,
$$\prod_{s\in S} a_s = \prod_{w=1}^{n} \prod_{s\in S_w} a_s.$$

- **Fubini's theorem (interchanging the order of multiplication):** Let X and Y be two finite sets. Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x, y) \in X \times Y$. Then,
$$\prod_{x\in X} \prod_{y\in Y} a_{(x,y)} = \prod_{(x,y)\in X\times Y} a_{(x,y)} = \prod_{y\in Y} \prod_{x\in X} a_{(x,y)}.$$

  In particular, if n and m are two elements of $\mathbb{N}$, and if $a_{(x,y)}$ is an element of $\mathbb{A}$ for each $(x, y) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$, then
$$\prod_{x=1}^{n} \prod_{y=1}^{m} a_{(x,y)} = \prod_{(x,y)\in\{1,2,\ldots,n\}\times\{1,2,\ldots,m\}} a_{(x,y)} = \prod_{y=1}^{m} \prod_{x=1}^{n} a_{(x,y)}.$$

- **Triangular Fubini's theorem I:** Let $n \in \mathbb{N}$. Let T_n be the set $\left\{ (x, y) \in \{1, 2, 3, \ldots\}^2 \mid x + y \leq n \right\}$. Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x, y) \in T_n$. Then,
$$\prod_{x=1}^{n} \prod_{y=1}^{n-x} a_{(x,y)} = \prod_{(x,y)\in T_n} a_{(x,y)} = \prod_{y=1}^{n} \prod_{x=1}^{n-y} a_{(x,y)}.$$

- **Triangular Fubini's theorem II:** Let $n \in \mathbb{N}$. Let Q_n be the set $\left\{ (x, y) \in \{1, 2, \ldots, n\}^2 \mid x \leq y \right\}$. Let $a_{(x,y)}$ be an element of $\mathbb{A}$ for each $(x, y) \in Q_n$. Then,
$$\prod_{x=1}^{n} \prod_{y=x}^{n} a_{(x,y)} = \prod_{(x,y)\in Q_n} a_{(x,y)} = \prod_{y=1}^{n} \prod_{x=1}^{y} a_{(x,y)}.$$

- **Fubini's theorem with a predicate:** Let X and Y be two finite sets. For every pair $(x, y) \in X \times Y$, let $\mathcal{A}(x, y)$ be a logical statement. For each $(x, y) \in X \times Y$ satisfying $\mathcal{A}(x, y)$, let $a_{(x,y)}$ be an element of $\mathbb{A}$. Then,
$$\prod_{x\in X} \prod_{\substack{y\in Y;\\ \mathcal{A}(x,y)}} a_{(x,y)} = \prod_{\substack{(x,y)\in X\times Y;\\ \mathcal{A}(x,y)}} a_{(x,y)} = \prod_{y\in Y} \prod_{\substack{x\in X;\\ \mathcal{A}(x,y)}} a_{(x,y)}.$$

- **Interchange of predicates:** Let S be a finite set. For every $s \in S$, let $\mathcal{A}(s)$ and $\mathcal{B}(s)$ be two equivalent logical statements. ("Equivalent" means that $\mathcal{A}(s)$ holds if and only if $\mathcal{B}(s)$ holds.) Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\prod_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s = \prod_{\substack{s \in S; \\ \mathcal{B}(s)}} a_s.$$

- **Substituting the index I with a predicate:** Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_t be an element of $\mathbb{A}$ for each $t \in T$. For every $t \in T$, let $\mathcal{A}(t)$ be a logical statement. Then,

$$\prod_{\substack{t \in T; \\ \mathcal{A}(t)}} a_t = \prod_{\substack{s \in S; \\ \mathcal{A}(f(s))}} a_{f(s)}.$$

1.5. Polynomials: a precise definition

As I have already mentioned in the above list of prerequisites, the notion of polynomials (in one and in several indeterminates) will be occasionally used in these notes. Most likely, the reader already has at least a vague understanding of this notion (e.g., from high school); this vague understanding is probably sufficient for reading most of these notes. But polynomials are one of the most important notions in algebra (if not to say in mathematics), and the reader will likely encounter them over and over; sooner or later, it will happen that the vague understanding is not sufficient and some subtleties do matter. For that reason, anyone serious about doing abstract algebra should know a complete and correct definition of polynomials and have some experience working with it. I shall not give a complete definition of the most general notion of polynomials in these notes, but I will comment on some of the subtleties and define an important special case (that of polynomials in one variable with rational coefficients) in the present section. A reader is probably best advised to skip this section on their first read.

It is not easy to find a good (formal and sufficiently general) treatment of polynomials in textbooks. Various authors tend to skimp on subtleties and technical points such as the notion of an "indeterminate", or the precise meaning of "formal expression" in the slogan "a polynomial is a formal expression" (the best texts do not use this vague slogan at all), or the definition of the degree of the zero polynomial, or the difference between regarding polynomials as sequences (which is the classical viewpoint and particularly useful for polynomials in one variable) and regarding polynomials as elements of a monoid ring (which is important in the case of several variables, since it allows us to regard the polynomial rings $\mathbb{Q}[X]$ and $\mathbb{Q}[Y]$ as two distinct subrings of $\mathbb{Q}[X,Y]$). They also tend to take some questionable shortcuts, such as defining polynomials in n variables (by induction over n) as one-variable polynomials over the ring of $(n-1)$-variable polynomials (this shortcut has several shortcomings, such as making the symmetric role of the n variables opaque, and functioning only for finitely many variables).

More often than not, the polynomials we will be using will be polynomials in one variable. These are usually handled well in good books on abstract algebra – e.g., in [Walker87, §4.5], in [Hunger14, Appendix G], in [Hunger03, Chapter III, §5], in [Rotman15, Chapter A-3], in [HofKun71, §4.1, §4.2] (although in [HofKun71, §4.1, §4.2], only polynomials over fields are studied, but the definition applies to commutative rings mutatis mutandis), in [AmaEsc05, §8], and in [BirMac99, Chapter III, §6]. Most of these treatments rely on the notion of a *commutative ring*, which is not difficult but somewhat abstract (I shall introduce it below in Section 6.1).

Let me give a brief survey of the notion of univariate polynomials (i.e., polynomials in one variable). I shall define them as sequences. For the sake of simplicity, I shall only talk of polynomials with rational coefficients. Similarly, one can define polynomials with integer coefficients, with real coefficients, or with complex coefficients; of course, one then has to replace each "$\mathbb{Q}$" by a "$\mathbb{Z}$", an "$\mathbb{R}$" or a "$\mathbb{C}$".

The rough idea behind the definition of a polynomial is that a polynomial with rational coefficients should be a "formal expression" which is built out of rational numbers, an "indeterminate" X as well as addition, subtraction and multiplication signs, such as $X^4 - 27X + \dfrac{3}{2}$ or $-X^3 + 2X + 1$ or $\dfrac{1}{3}(X-3) \cdot X^2$ or $X^4 + 7X^3(X-2)$ or -15. We have not explicitly allowed powers, but we understand X^n to mean the product $\underbrace{XX \cdots X}_{n \text{ times}}$ (which is 1 when $n = 0$). Notice that division is not allowed, so we cannot get $\dfrac{X}{X+1}$ (but we can get $\dfrac{3}{2}X$, because $\dfrac{3}{2}$ is a rational number). Notice also that a polynomial can be a single rational number, since we never said that X must necessarily be used; for instance, -15 and 0 are polynomials.

This is, of course, not a valid definition. One problem with it that it does not explain what a "formal expression" is. For starters, we want an expression that is well-defined – i.e., into that we can substitute a rational number for X and obtain a valid term. For example, $X - + \cdot 5$ is not well-defined, so it does not fit our bill; neither is the "empty expression". Furthermore, when do we want two "formal expressions" to be viewed as one and the same polynomial? Do we want to equate $X(X+2)$ with $X^2 + 2X$? Do we want to equate $0X^3 + 2X + 1$ with $2X + 1$? The answer is "yes" both times, but a general rule is not easy to give if we keep talking of "formal expressions".

We *could* define two polynomials $p(X)$ and $q(X)$ to be equal if and only if, for every number $\alpha \in \mathbb{Q}$, the values $p(\alpha)$ and $q(\alpha)$ (obtained by substituting α for X in p and in q, respectively) are equal. This would be tantamount to treating polynomials as *functions*: it would mean that we identify a polynomial $p(X)$ with the function $\mathbb{Q} \to \mathbb{Q}$, $\alpha \mapsto p(\alpha)$. Such a definition would work well as long as we would do only rather basic things with it[25], but as soon as we would try to go deeper, we would encounter technical issues which would make it inadequate and painful[26]. Also, if we equated polynomials with the functions they describe, then

[25] And some authors, such as Axler in [Axler15, Chapter 4], do use this definition.

[26] Here are three of these issues:

we would waste the word "polynomial" on a concept (a function described by a polynomial) that already has a word for it (namely, *polynomial function*).

The preceding paragraphs indicate that it is worth defining "polynomials" in a way that, on the one hand, conveys the idea that they are more "formal expressions" than "functions", but on the other hand, is less nebulous than "formal expression". Here is one such definition:

Definition 1.7. (a) A *univariate polynomial with rational coefficients* means a sequence $(p_0, p_1, p_2, \ldots) \in \mathbb{Q}^\infty$ of elements of $\mathbb{Q}$ such that

$$\text{all but finitely many } k \in \mathbb{N} \text{ satisfy } p_k = 0. \qquad (41)$$

- One of the strengths of polynomials is that we can evaluate them not only at numbers, but also at many other things, e.g., at square matrices: Evaluating the polynomial $X^2 - 3X$ at the square matrix $\begin{pmatrix} 1 & 3 \\ -1 & 2 \end{pmatrix}$ gives $\begin{pmatrix} 1 & 3 \\ -1 & 2 \end{pmatrix}^2 - 3\begin{pmatrix} 1 & 3 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} -5 & 0 \\ 0 & -5 \end{pmatrix}$. However, a function must have a well-defined domain, and does not make sense outside of this domain. So, if the polynomial $X^2 - 3X$ is regarded as the function $\mathbb{Q} \to \mathbb{Q}$, $\alpha \mapsto \alpha^2 - 3\alpha$, then it makes no sense to evaluate this polynomial at the matrix $\begin{pmatrix} 1 & 3 \\ -1 & 2 \end{pmatrix}$, just because this matrix does not lie in the domain $\mathbb{Q}$ of the function. We could, of course, extend the domain of the function to (say) the set of square matrices over $\mathbb{Q}$, but then we would still have the same problem with other things that we want to evaluate polynomials at. At some point we want to be able to evaluate polynomials at functions and at other polynomials, and if we would try to achieve this by extending the domain, we would have to do this over and over, because each time we extend the domain, we get even more polynomials to evaluate our polynomials at; thus, the definition would be eternally "hunting its own tail"! (We could resolve this difficulty by defining polynomials as *natural transformations* in the sense of category theory. I do not want to even go into this definition here, as it would take several pages to properly introduce. At this point, it is not worth the hassle.)
- Let $p(X)$ be a polynomial with real coefficients. Then, it should be obvious that $p(X)$ can also be viewed as a polynomial with complex coefficients: For instance, if $p(X)$ was defined as $3X + \frac{7}{2}X(X-1)$, then we can view the numbers 3, $\frac{7}{2}$ and -1 appearing in its definition as complex numbers, and thus get a polynomial with complex coefficients. But wait! What if two polynomials $p(X)$ and $q(X)$ are equal when viewed as polynomials with real coefficients, but become distinct when viewed as polynomials with complex coefficients (because when we view them as polynomials with complex coefficients, their domains grow larger to include complex numbers, and a new complex α might perhaps no longer satisfy $p(\alpha) = q(\alpha)$)? This does not actually happen, but ruling this out is not obvious if you regard polynomials as functions.
- (This requires some familiarity with finite fields:) Treating polynomials as functions works reasonably well for polynomials with integer, rational, real and complex coefficients (as long as one is not too demanding). But we will eventually want to consider polynomials with coefficients in any arbitrary commutative ring $\mathbb{K}$. An example for a commutative ring $\mathbb{K}$ is the finite field $\mathbb{F}_p$ with p elements, where p is a prime. (This finite field $\mathbb{F}_p$ is better known as the ring of integers modulo p.) If we define polynomials with coefficients in $\mathbb{F}_p$ as functions $\mathbb{F}_p \to \mathbb{F}_p$, then we really run into problems; for example, the polynomials X and X^p over this field become identical as functions!

Here, the phrase "all but finitely many $k \in \mathbb{N}$ satisfy $p_k = 0$" means "there exists some finite subset J of $\mathbb{N}$ such that every $k \in \mathbb{N} \setminus J$ satisfies $p_k = 0$". (See Definition 5.17 for the general definition of "all but finitely many", and Section 5.4 for some practice with this concept.) More concretely, the condition (41) can be rewritten as follows: The sequence $(p_0, p_1, p_2, \ldots)$ contains only zeroes from some point on (i.e., there exists some $N \in \mathbb{N}$ such that $p_N = p_{N+1} = p_{N+2} = \cdots = 0$).

For the remainder of this definition, "univariate polynomial with rational coefficients" will be abbreviated as "polynomial".

For example, the sequences $(0,0,0,\ldots)$, $(1,3,5,0,0,0,\ldots)$, $\left(4,0,-\dfrac{2}{3},5,0,0,0,\ldots\right)$, $\left(0,-1,\dfrac{1}{2},0,0,0,\ldots\right)$ (where the "$\ldots$" stand for infinitely many zeroes) are polynomials, but the sequence $(1,1,1,\ldots)$ (where the "$\ldots$" stands for infinitely many 1's) is not (since it does not satisfy (41)).

So we have defined a polynomial as an infinite sequence of rational numbers with a certain property. So far, this does not seem to reflect any intuition of polynomials as "formal expressions". However, we shall soon (namely, in Definition 1.7 **(j)**) identify the polynomial $(p_0, p_1, p_2, \ldots) \in \mathbb{Q}^\infty$ with the "formal expression" $p_0 + p_1 X + p_2 X^2 + \cdots$ (this is an infinite sum, but due to (41) all but its first few terms are 0 and thus can be neglected). For instance, the polynomial $(1,3,5,0,0,0,\ldots)$ will be identified with the "formal expression" $1 + 3X + 5X^2 + 0X^3 + 0X^4 + 0X^5 + \cdots = 1 + 3X + 5X^2$. Of course, we cannot do this identification right now, since we do not have a reasonable definition of X.

(b) We let $\mathbb{Q}[X]$ denote the set of all univariate polynomials with rational coefficients. Given a polynomial $p = (p_0, p_1, p_2, \ldots) \in \mathbb{Q}[X]$, we denote the numbers $p_0, p_1, p_2, \ldots$ as the *coefficients* of p. More precisely, for every $i \in \mathbb{N}$, we shall refer to p_i as the *i-th coefficient* of p. (Do not forget that we are counting from 0 here: any polynomial "begins" with its 0-th coefficient.) The 0-th coefficient of p is also known as the *constant term* of p.

Instead of "the i-th coefficient of p", we often also say "the *coefficient before X^i of p*" or "the *coefficient of X^i in p*".

Thus, any polynomial $p \in \mathbb{Q}[X]$ is the sequence of its coefficients.

(c) We denote the polynomial $(0,0,0,\ldots) \in \mathbb{Q}[X]$ by $\mathbf{0}$. We will also write 0 for it when no confusion with the number 0 is possible. The polynomial $\mathbf{0}$ is called the *zero polynomial*. A polynomial $p \in \mathbb{Q}[X]$ is said to be *nonzero* if $p \neq \mathbf{0}$.

(d) We denote the polynomial $(1,0,0,0,\ldots) \in \mathbb{Q}[X]$ by $\mathbf{1}$. We will also write 1 for it when no confusion with the number 1 is possible.

(e) For any $\lambda \in \mathbb{Q}$, we denote the polynomial $(\lambda,0,0,0,\ldots) \in \mathbb{Q}[X]$ by $\operatorname{const}\lambda$. We call it the *constant polynomial with value λ*. It is often useful to identify $\lambda \in \mathbb{Q}$ with $\operatorname{const}\lambda \in \mathbb{Q}[X]$. Notice that $\mathbf{0} = \operatorname{const} 0$ and $\mathbf{1} = \operatorname{const} 1$.

(f) Now, let us define the sum, the difference and the product of two polynomials. Indeed, let $a = (a_0, a_1, a_2, \ldots) \in \mathbb{Q}[X]$ and $b = (b_0, b_1, b_2, \ldots) \in \mathbb{Q}[X]$ be two polynomials. Then, we define three polynomials $a + b$, $a - b$ and $a \cdot b$ in

$\mathbb{Q}[X]$ by

$$a+b = (a_0+b_0, a_1+b_1, a_2+b_2, \ldots);$$
$$a-b = (a_0-b_0, a_1-b_1, a_2-b_2, \ldots);$$
$$a \cdot b = (c_0, c_1, c_2, \ldots),$$

where

$$c_k = \sum_{i=0}^{k} a_i b_{k-i} \qquad \text{for every } k \in \mathbb{N}.$$

We call $a+b$ the *sum* of a and b; we call $a-b$ the *difference* of a and b; we call $a \cdot b$ the *product* of a and b. We abbreviate $a \cdot b$ by ab, and we abbreviate $\mathbf{0} - a$ by $-a$.

For example,

$$(1,2,2,0,0,\ldots) + (3,0,-1,0,0,0,\ldots) = (4,2,1,0,0,0,\ldots);$$
$$(1,2,2,0,0,\ldots) - (3,0,-1,0,0,0,\ldots) = (-2,2,3,0,0,0,\ldots);$$
$$(1,2,2,0,0,\ldots) \cdot (3,0,-1,0,0,0,\ldots) = (3,6,5,-2,-2,0,0,0,\ldots).$$

The definition of $a+b$ essentially says that "polynomials are added coefficientwise" (i.e., in order to obtain the sum of two polynomials a and b, it suffices to add each coefficient of a to the corresponding coefficient of b). Similarly, the definition of $a-b$ says the same thing about subtraction. The definition of $a \cdot b$ is more surprising. However, it loses its mystique when we identify the polynomials a and b with the "formal expressions" $a_0 + a_1X + a_2X^2 + \cdots$ and $b_0 + b_1X + b_2X^2 + \cdots$ (although, at this point, we do not know what these expressions really mean); indeed, it simply says that

$$\left(a_0 + a_1X + a_2X^2 + \cdots\right)\left(b_0 + b_1X + b_2X^2 + \cdots\right) = c_0 + c_1X + c_2X^2 + \cdots,$$

where $c_k = \sum\limits_{i=0}^{k} a_i b_{k-i}$ for every $k \in \mathbb{N}$. This is precisely what one would expect, because if you expand $\left(a_0 + a_1X + a_2X^2 + \cdots\right)\left(b_0 + b_1X + b_2X^2 + \cdots\right)$ using the distributive law and collect equal powers of X, then you get precisely $c_0 + c_1X + c_2X^2 + \cdots$. Thus, the definition of $a \cdot b$ has been tailored to make the distributive law hold.

(By the way, why is $a \cdot b$ a polynomial? That is, why does it satisfy (41) ? The proof is easy, but we omit it.)

Addition, subtraction and multiplication of polynomials satisfy some of the same rules as addition, subtraction and multiplication of numbers. For example, the commutative laws $a+b = b+a$ and $ab = ba$ are valid for polynomials just as they are for numbers; the same holds for the associative laws $(a+b)+c = a+(b+c)$ and $(ab)c = a(bc)$ and the distributive laws $(a+b)c = ac+bc$ and $a(b+c) = ab+ac$. Moreover, each polynomial a satisfies $a + \mathbf{0} = \mathbf{0} + a = a$ and $a \cdot \mathbf{0} = \mathbf{0} \cdot a = \mathbf{0}$ and $a \cdot \mathbf{1} = \mathbf{1} \cdot a = a$ and $a + (-a) = (-a) + a = \mathbf{0}$.

Using the notations of Definition 6.2, we can summarize this as follows: The set $\mathbb{Q}[X]$, endowed with the operations $+$ and $\cdot$ just defined, and with the elements $\mathbf{0}$ and $\mathbf{1}$, is a commutative ring. It is called the *(univariate) polynomial ring over* $\mathbb{Q}$.

(g) Let $a = (a_0, a_1, a_2, \ldots) \in \mathbb{Q}[X]$ and $\lambda \in \mathbb{Q}$. Then, λa denotes the polynomial $(\lambda a_0, \lambda a_1, \lambda a_2, \ldots) \in \mathbb{Q}[X]$. (This equals the polynomial $(\operatorname{const} \lambda) \cdot a$; thus, identifying λ with $\operatorname{const} \lambda$ does not cause any inconsistencies here.)

(h) If $p = (p_0, p_1, p_2, \ldots) \in \mathbb{Q}[X]$ is a nonzero polynomial, then the *degree* of p is defined to be the maximum $i \in \mathbb{N}$ satisfying $p_i \neq 0$. If $p \in \mathbb{Q}[X]$ is the zero polynomial, then the degree of p is defined to be $-\infty$. (Here, $-\infty$ is just a fancy symbol, not a number.) For example, $\deg(0, 4, 0, -1, 0, 0, 0, \ldots) = 3$.

(i) If $a = (a_0, a_1, a_2, \ldots) \in \mathbb{Q}[X]$ and $n \in \mathbb{N}$, then a polynomial $a^n \in \mathbb{Q}[X]$ is defined to be the product $\underbrace{aa \cdots a}_{n \text{ times}}$. (This is understood to be $\mathbf{1}$ when $n = 0$. In general, an empty product of polynomials is always understood to be $\mathbf{1}$.)

(j) We let X denote the polynomial $(0, 1, 0, 0, 0, \ldots) \in \mathbb{Q}[X]$. (This is the polynomial whose 1-st coefficient is 1 and whose other coefficients are 0.) This polynomial is called the *indeterminate* of $\mathbb{Q}[X]$. It is easy to see that, for any $n \in \mathbb{N}$, we have

$$X^n = \left(\underbrace{0, 0, \ldots, 0}_{n \text{ zeroes}}, 1, 0, 0, 0, \ldots \right).$$

This polynomial X finally provides an answer to the questions "what is an indeterminate" and "what is a formal expression". Namely, let $(p_0, p_1, p_2, \ldots) \in \mathbb{Q}[X]$ be any polynomial. Then, the sum $p_0 + p_1 X + p_2 X^2 + \cdots$ is well-defined (it is an infinite sum, but due to (41) it has only finitely many nonzero addends), and it is easy to see that this sum equals $(p_0, p_1, p_2, \ldots)$. Thus,

$$(p_0, p_1, p_2, \ldots) = p_0 + p_1 X + p_2 X^2 + \cdots \qquad \text{for every } (p_0, p_1, p_2, \ldots) \in \mathbb{Q}[X].$$

This finally allows us to write a polynomial $(p_0, p_1, p_2, \ldots)$ as a sum $p_0 + p_1 X + p_2 X^2 + \cdots$ while remaining honest; the sum $p_0 + p_1 X + p_2 X^2 + \cdots$ is no longer a "formal expression" of unclear meaning, nor a function, but it is just an alternative way to write the sequence $(p_0, p_1, p_2, \ldots)$. So, at last, our notion of a polynomial resembles the intuitive notion of a polynomial!

Of course, we can write polynomials as finite sums as well. Indeed, if $(p_0, p_1, p_2, \ldots) \in \mathbb{Q}[X]$ is a polynomial and N is a nonnegative integer such that every $n > N$ satisfies $p_n = 0$, then

$$(p_0, p_1, p_2, \ldots) = p_0 + p_1 X + p_2 X^2 + \cdots = p_0 + p_1 X + \cdots + p_N X^N$$

(because addends can be discarded when they are 0). For example,

$$(4, 1, 0, 0, 0, \ldots) = 4 + 1X = 4 + X \qquad \text{and}$$
$$\left(\frac{1}{2}, 0, \frac{1}{3}, 0, 0, 0, \ldots \right) = \frac{1}{2} + 0X + \frac{1}{3} X^2 = \frac{1}{2} + \frac{1}{3} X^2.$$

(k) For our definition of polynomials to be fully compatible with our intuition, we are missing only one more thing: a way to evaluate a polynomial at a number, or some other object (e.g., another polynomial or a function). This is easy: Let $p = (p_0, p_1, p_2, \ldots) \in \mathbb{Q}[X]$ be a polynomial, and let $\alpha \in \mathbb{Q}$. Then, $p(\alpha)$ means the number $p_0 + p_1\alpha + p_2\alpha^2 + \cdots \in \mathbb{Q}$. (Again, the infinite sum $p_0 + p_1\alpha + p_2\alpha^2 + \cdots$ makes sense because of (41).) Similarly, we can define $p(\alpha)$ when $\alpha \in \mathbb{R}$ (but in this case, $p(\alpha)$ will be an element of $\mathbb{R}$) or when $\alpha \in \mathbb{C}$ (in this case, $p(\alpha) \in \mathbb{C}$) or when α is a square matrix with rational entries (in this case, $p(\alpha)$ will also be such a matrix) or when α is another polynomial (in this case, $p(\alpha)$ is such a polynomial as well).

For example, if $p = (1, -2, 0, 3, 0, 0, 0, \ldots) = 1 - 2X + 3X^3$, then $p(\alpha) = 1 - 2\alpha + 3\alpha^3$ for every α.

The map $\mathbb{Q} \to \mathbb{Q},\ \alpha \mapsto p(\alpha)$ is called the *polynomial function described by* p. As we said above, this function is not p, and it is not a good idea to equate it with p.

If α is a number (or a square matrix, or another polynomial), then $p(\alpha)$ is called the result of *evaluating* p *at* $X = \alpha$ (or, simply, evaluating p at α), or the result of *substituting* α *for* X *in* p. This notation, of course, reminds of functions; nevertheless, (as we already said a few times) p is **not a function**.

Probably the simplest three cases of evaluation are the following ones:

- We have $p(0) = p_0 + p_1 0^1 + p_2 0^2 + \cdots = p_0$. In other words, evaluating p at $X = 0$ yields the constant term of p.
- We have $p(1) = p_0 + p_1 1^1 + p_2 1^2 + \cdots = p_0 + p_1 + p_2 + \cdots$. In other words, evaluating p at $X = 1$ yields the sum of all coefficients of p.
- We have $p(X) = p_0 + p_1 X^1 + p_2 X^2 + \cdots = p_0 + p_1 X + p_2 X^2 + \cdots = p$. In other words, evaluating p at $X = X$ yields p itself. This allows us to write $p(X)$ for p. Many authors do so, just in order to stress that p is a polynomial and that the indeterminate is called X. It should be kept in mind that X is **not a variable** (just as p is **not a function**); it is the (fixed!) sequence $(0, 1, 0, 0, 0, \ldots) \in \mathbb{Q}[X]$ which serves as the indeterminate for polynomials in $\mathbb{Q}[X]$.

(l) Often, one wants (or is required) to give an indeterminate a name other than X. (For instance, instead of polynomials with rational coefficients, we could be considering polynomials whose coefficients themselves are polynomials in $\mathbb{Q}[X]$; and then, we would not be allowed to use the letter X for the "new" indeterminate anymore, as it already means the indeterminate of $\mathbb{Q}[X]$!) This can be done, and the rules are the following: Any letter (that does not already have a meaning) can be used to denote the indeterminate; but then, the set of all polynomials has to be renamed as $\mathbb{Q}[\eta]$, where η is this letter. For instance, if we want to denote the indeterminate as x, then we have to denote the set by $\mathbb{Q}[x]$.

It is furthermore convenient to regard the sets $\mathbb{Q}[\eta]$ for different letters η as distinct. Thus, for example, the polynomial $3X^2 + 1$ is not the same as the polynomial $3Y^2 + 1$. (The reason for doing so is that one sometimes wishes to view both of these polynomials as polynomials in the two variables X and Y.) Formally speaking, this means that we should define a polynomial in $\mathbb{Q}[\eta]$ to be not just a sequence $(p_0, p_1, p_2, \ldots)$ of rational numbers, but actually a pair $((p_0, p_1, p_2, \ldots), \text{“}\eta\text{”})$ of a sequence of rational numbers and the letter η. (Here, “η” really means the letter η, not the sequence $(0, 1, 0, 0, 0, \ldots)$.) This is, of course, a very technical point which is of little relevance to most of mathematics; it becomes important when one tries to implement polynomials in a programming language.

(m) As already explained, we can replace $\mathbb{Q}$ by $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$ or any other commutative ring $\mathbb{K}$ in the above definition. (See Definition 6.2 for the definition of a commutative ring.) When $\mathbb{Q}$ is replaced by a commutative ring $\mathbb{K}$, the notion of “univariate polynomials with rational coefficients” becomes “univariate polynomials with coefficients in $\mathbb{K}$” (also known as “univariate polynomials over $\mathbb{K}$”), and the set of such polynomials is denoted by $\mathbb{K}[X]$ rather than $\mathbb{Q}[X]$.

So much for univariate polynomials.

Polynomials in multiple variables are (in my opinion) treated the best in [Lang02, Chapter II, §3], where they are introduced as elements of a monoid ring. However, this treatment is rather abstract and uses a good deal of algebraic language[27]. The treatments in [Walker87, §4.5], in [Rotman15, Chapter A-3] and in [BirMac99, Chapter IV, §4] use the above-mentioned recursive shortcut that makes them inferior (in my opinion). A neat (and rather elementary) treatment of polynomials in n variables (for finite n) can be found in [Hunger03, Chapter III, §5], in [Loehr11, §7.16], in [GalQua18, §30.2] and in [AmaEsc05, §I.8]; it generalizes the viewpoint we used in Definition 1.7 for univariate polynomials above[28].

2. A closer look at induction

In this chapter, we shall recall several versions of the *induction principle* (the principle of mathematical induction) and provide examples for their use. We assume that the reader is at least somewhat familiar with mathematical induction[29]; we shall present some nonstandard examples of its use (including a proof of the legitimacy of the definition of a sum $\sum\limits_{s \in S} a_s$ given in Section 1.4).

[27] Also, the book [Lang02] is notorious for its unpolished writing; it is best read with Bergman's companion [Bergma15] at hand.

[28] You are reading right: The analysis textbook [AmaEsc05] is one of the few sources I am aware of to define the (algebraic!) notion of polynomials precisely and well.

[29] If not, introductions can be found in [LeLeMe16, Chapter 5], [Day16], [Vellem06, Chapter 6], [Hammac15, Chapter 10], [Vorobi02] and various other sources.

2.1. Standard induction

2.1.1. The Principle of Mathematical Induction

We first recall the classical principle of mathematical induction[30]:

Theorem 2.1. For each $n \in \mathbb{N}$, let $\mathcal{A}(n)$ be a logical statement.
Assume the following:

Assumption 1: The statement $\mathcal{A}(0)$ holds.

Assumption 2: If $m \in \mathbb{N}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \mathbb{N}$.

Theorem 2.1 is commonly taken to be one of the axioms of mathematics (the "axiom of induction"), or (in type theory) as part of the definition of $\mathbb{N}$. Intuitively, Theorem 2.1 should be obvious: For example, if you want to prove (under the assumptions of Theorem 2.1) that $\mathcal{A}(4)$ holds, you can argue as follows:

- By Assumption 1, the statement $\mathcal{A}(0)$ holds.
- Thus, by Assumption 2 (applied to $m = 0$), the statement $\mathcal{A}(1)$ holds.
- Thus, by Assumption 2 (applied to $m = 1$), the statement $\mathcal{A}(2)$ holds.
- Thus, by Assumption 2 (applied to $m = 2$), the statement $\mathcal{A}(3)$ holds.
- Thus, by Assumption 2 (applied to $m = 3$), the statement $\mathcal{A}(4)$ holds.

A similar (but longer) argument shows that the statement $\mathcal{A}(5)$ holds. Likewise, you can show that the statement $\mathcal{A}(15)$ holds, if you have the patience to apply Assumption 2 a total of 15 times. It is thus not surprising that $\mathcal{A}(n)$ holds for each $n \in \mathbb{N}$; but if you don't assume Theorem 2.1 as an axiom, you would need to write down a different proof for each value of n (which becomes the longer the larger n is), and thus would never reach the general result (i.e., that $\mathcal{A}(n)$ holds for **each** $n \in \mathbb{N}$), because you cannot write down infinitely many proofs. What Theorem 2.1 does is, roughly speaking, to apply Assumption 2 for you as many times as it is needed for each $n \in \mathbb{N}$.

(Authors of textbooks like to visualize Theorem 2.1 by envisioning an infinite sequence of dominos (numbered $0, 1, 2, \ldots$) placed in row, sufficiently close to each other that if domino m falls, then domino $m+1$ will also fall. Now, assume that you kick domino 0 over. What Theorem 2.1 then says is that each domino will fall. See, e.g., [Hammac15, Chapter 10] for a detailed explanation of this metaphor.

[30] Keep in mind that $\mathbb{N}$ means the set $\{0, 1, 2, \ldots\}$ for us.

Here is another metaphor for Theorem 2.1: Assume that there is a virus that infects nonnegative integers. Once it has infected some $m \in \mathbb{N}$, it will soon spread to $m + 1$ as well. Now, assume that 0 gets infected. Then, Theorem 2.1 says that each $n \in \mathbb{N}$ will eventually be infected.)

Theorem 2.1 is called the *principle of induction* or *principle of complete induction* or *principle of mathematical induction*, and we shall also call it *principle of standard induction* in order to distinguish it from several variant "principles of induction" that we will see later. Proofs that use this principle are called *proofs by induction* or *induction proofs*. Usually, in such proofs, we don't explicitly cite Theorem 2.1, but instead say certain words that signal that Theorem 2.1 is being applied and that (ideally) also indicate what statements $\mathcal{A}(n)$ it is being applied to[31]. However, for our very first example of a proof by induction, we are going to use Theorem 2.1 explicitly. We shall show the following fact:

Proposition 2.2. Let q and d be two real numbers such that $q \neq 1$. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of real numbers. Assume that

$$a_{n+1} = qa_n + d \qquad \text{for each } n \in \mathbb{N}. \tag{42}$$

Then,

$$a_n = q^n a_0 + \frac{q^n - 1}{q - 1} d \qquad \text{for each } n \in \mathbb{N}. \tag{43}$$

Proof of Proposition 2.2. For each $n \in \mathbb{N}$, we let $\mathcal{A}(n)$ be the statement $\left(a_n = q^n a_0 + \frac{q^n - 1}{q - 1} d\right)$. Thus, our goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \mathbb{N}$.

We first notice that the statement $\mathcal{A}(0)$ holds[32].

Now, we claim that

$$\text{if } m \in \mathbb{N} \text{ is such that } \mathcal{A}(m) \text{ holds, then } \mathcal{A}(m+1) \text{ also holds.} \tag{44}$$

[*Proof of (44):* Let $m \in \mathbb{N}$ be such that $\mathcal{A}(m)$ holds. We must show that $\mathcal{A}(m+1)$ also holds.

[31] We will explain this in Convention 2.3 below.

[32] *Proof.* This is easy to verify: We have $q^0 = 1$, thus $q^0 - 1 = 0$, and therefore $\frac{q^0 - 1}{q - 1} = \frac{0}{q - 1} = 0$. Now,

$$\underbrace{q^0}_{=1} a_0 + \underbrace{\frac{q^0 - 1}{q - 1}}_{=0} d = 1a_0 + 0d = a_0,$$

so that $a_0 = q^0 a_0 + \frac{q^0 - 1}{q - 1} d$. But this is precisely the statement $\mathcal{A}(0)$ (since $\mathcal{A}(0)$ is defined to be the statement $\left(a_0 = q^0 a_0 + \frac{q^0 - 1}{q - 1} d\right)$). Hence, the statement $\mathcal{A}(0)$ holds.

We have assumed that $\mathcal{A}(m)$ holds. In other words, $a_m = q^m a_0 + \dfrac{q^m - 1}{q-1} d$ holds[33]. Now, (42) (applied to $n = m$) yields

$$\begin{aligned} a_{m+1} &= q \underbrace{a_m}_{= q^m a_0 + \frac{q^m-1}{q-1}d} + d = q\left(q^m a_0 + \frac{q^m-1}{q-1}d\right) + d \\ &= \underbrace{qq^m}_{=q^{m+1}} a_0 + \underbrace{q \cdot \frac{q^m-1}{q-1}d + d}_{=\left(q\cdot \frac{q^m-1}{q-1}+1\right)d} \\ &= q^{m+1}a_0 + \underbrace{\left(q\cdot\frac{q^m-1}{q-1}+1\right)}_{\substack{=\frac{q(q^m-1)+(q-1)}{q-1}=\frac{q^{m+1}-1}{q-1} \\ (\text{since } q(q^m-1)+(q-1)=qq^m-q+q-1=qq^m-1=q^{m+1}-1)}} d \\ &= q^{m+1}a_0 + \frac{q^{m+1}-1}{q-1}d. \end{aligned}$$

So we have shown that $a_{m+1} = q^{m+1}a_0 + \dfrac{q^{m+1}-1}{q-1}d$. But this is precisely the statement $\mathcal{A}(m+1)$ [34]. Thus, the statement $\mathcal{A}(m+1)$ holds.

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{N}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. This proves (44).]

Now, both assumptions of Theorem 2.1 are satisfied (indeed, Assumption 1 holds because the statement $\mathcal{A}(0)$ holds, whereas Assumption 2 holds because of (44)). Thus, Theorem 2.1 shows that $\mathcal{A}(n)$ holds for each $n \in \mathbb{N}$. In other words, $a_n = q^n a_0 + \dfrac{q^n-1}{q-1}d$ holds for each $n \in \mathbb{N}$ (since $\mathcal{A}(n)$ is the statement $\left(a_n = q^n a_0 + \dfrac{q^n-1}{q-1}d\right)$). This proves Proposition 2.2. □

2.1.2. Conventions for writing induction proofs

Now, let us introduce some standard language that is commonly used in proofs by induction:

---

[33]because $\mathcal{A}(m)$ is defined to be the statement $\left(a_m = q^m a_0 + \dfrac{q^m-1}{q-1}d\right)$

[34]because $\mathcal{A}(m+1)$ is defined to be the statement $\left(a_{m+1} = q^{m+1}a_0 + \dfrac{q^{m+1}-1}{q-1}d\right)$

Convention 2.3. For each $n \in \mathbb{N}$, let $\mathcal{A}(n)$ be a logical statement. Assume that you want to prove that $\mathcal{A}(n)$ holds for each $n \in \mathbb{N}$.

Theorem 2.1 offers the following strategy for proving this: First show that Assumption 1 of Theorem 2.1 is satisfied; then, show that Assumption 2 of Theorem 2.1 is satisfied; then, Theorem 2.1 automatically completes your proof.

A proof that follows this strategy is called a *proof by induction on n* (or *proof by induction over n*) or (less precisely) an *inductive proof*. When you follow this strategy, you say that you are *inducting on n* (or *over n*). The proof that Assumption 1 is satisfied is called the *induction base* (or *base case*) of the proof. The proof that Assumption 2 is satisfied is called the *induction step* of the proof.

In order to prove that Assumption 2 is satisfied, you will usually want to fix an $m \in \mathbb{N}$ such that $\mathcal{A}(m)$ holds, and then prove that $\mathcal{A}(m+1)$ holds. In other words, you will usually want to fix $m \in \mathbb{N}$, assume that $\mathcal{A}(m)$ holds, and then prove that $\mathcal{A}(m+1)$ holds. When doing so, it is common to refer to the assumption that $\mathcal{A}(m)$ holds as the *induction hypothesis* (or *induction assumption*).

Using this language, we can rewrite our above proof of Proposition 2.2 as follows:

Proof of Proposition 2.2 (second version). For each $n \in \mathbb{N}$, we let $\mathcal{A}(n)$ be the statement $\left(a_n = q^n a_0 + \dfrac{q^n - 1}{q - 1} d\right)$. Thus, our goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \mathbb{N}$.

We shall prove this by induction on n:

Induction base: We have $q^0 = 1$, thus $q^0 - 1 = 0$, and therefore $\dfrac{q^0 - 1}{q - 1} = \dfrac{0}{q - 1} = 0$. Now,

$$\underbrace{q^0}_{=1} a_0 + \underbrace{\frac{q^0 - 1}{q - 1}}_{=0} d = 1a_0 + 0d = a_0,$$

so that $a_0 = q^0 a_0 + \dfrac{q^0 - 1}{q - 1} d$. But this is precisely the statement $\mathcal{A}(0)$ (since $\mathcal{A}(0)$ is defined to be the statement $\left(a_0 = q^0 a_0 + \dfrac{q^0 - 1}{q - 1} d\right)$). Hence, the statement $\mathcal{A}(0)$ holds. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that $\mathcal{A}(m)$ holds. We must show that $\mathcal{A}(m+1)$ also holds.

We have assumed that $\mathcal{A}(m)$ holds (this is our induction hypothesis). In other words, $a_m = q^m a_0 + \frac{q^m - 1}{q - 1} d$ holds[35]. Now, (42) (applied to $n = m$) yields

$$\begin{aligned} a_{m+1} &= q \underbrace{a_m}_{= q^m a_0 + \frac{q^m-1}{q-1} d} + d = q \left(q^m a_0 + \frac{q^m - 1}{q-1} d \right) + d \\ &= \underbrace{qq^m}_{= q^{m+1}} a_0 + \underbrace{q \cdot \frac{q^m - 1}{q-1} d + d}_{= \left(q \cdot \frac{q^m-1}{q-1} + 1 \right) d} \\ &= q^{m+1} a_0 + \underbrace{\left(q \cdot \frac{q^m - 1}{q - 1} + 1 \right)}_{\substack{= \frac{q(q^m-1)+(q-1)}{q-1} = \frac{q^{m+1}-1}{q-1} \\ \text{(since } q(q^m-1)+(q-1) = qq^m - q + q - 1 = qq^m - 1 = q^{m+1} - 1\text{)}}} d \\ &= q^{m+1} a_0 + \frac{q^{m+1} - 1}{q - 1} d. \end{aligned}$$

So we have shown that $a_{m+1} = q^{m+1} a_0 + \frac{q^{m+1} - 1}{q - 1} d$. But this is precisely the statement $\mathcal{A}(m+1)$ [36]. Thus, the statement $\mathcal{A}(m+1)$ holds.

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{N}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. This completes the induction step.

Thus, we have completed both the induction base and the induction step. Hence, by induction, we conclude that $\mathcal{A}(n)$ holds for each $n \in \mathbb{N}$. This proves Proposition 2.2. □

The proof we just gave still has a lot of "boilerplate" text. For example, we have explicitly defined the statement $\mathcal{A}(n)$, but it is not really necessary, since it is clear what this statement should be (viz., it should be the claim we are proving, without the "for each $n \in \mathbb{N}$" part). Allowing ourselves some imprecision, we could say this statement is simply (43). (This is a bit imprecise, because (43) contains the words "for each $n \in \mathbb{N}$", but it should be clear that we don't mean to include these words, since there can be no "for each $n \in \mathbb{N}$" in the statement $\mathcal{A}(n)$.) Furthermore, we don't need to write the sentence

> "Thus, we have completed both the induction base and the induction step"

[35] because $\mathcal{A}(m)$ is defined to be the statement $\left(a_m = q^m a_0 + \frac{q^m - 1}{q - 1} d \right)$

[36] because $\mathcal{A}(m+1)$ is defined to be the statement $\left(a_{m+1} = q^{m+1} a_0 + \frac{q^{m+1} - 1}{q - 1} d \right)$

before we declare our inductive proof to be finished; it is clear enough that we have completed them. We also can remove the following two sentences:

> "Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{N}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds.".

In fact, these sentences merely say that we have completed the induction step; they carry no other information (since the induction step always consists in fixing $m \in \mathbb{N}$ such that $\mathcal{A}(m)$ holds, and proving that $\mathcal{A}(m+1)$ also holds). So once we say that the induction step is completed, we don't need these sentences anymore.

So we can shorten our proof above a bit further:

Proof of Proposition 2.2 (third version). We shall prove (43) by induction on n:

Induction base: We have $q^0 = 1$, thus $q^0 - 1 = 0$, and therefore $\dfrac{q^0 - 1}{q - 1} = \dfrac{0}{q-1} = 0$. Now,

$$\underbrace{q^0}_{=1} a_0 + \underbrace{\frac{q^0 - 1}{q - 1}}_{=0} d = 1a_0 + 0d = a_0,$$

so that $a_0 = q^0 a_0 + \dfrac{q^0 - 1}{q - 1} d$. In other words, (43) holds for $n = 0$. [37] This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (43) holds for $n = m$. [38] We must show that (43) holds for $n = m + 1$. [39]

We have assumed that (43) holds for $n = m$. In other words, $a_m = q^m a_0 + \dfrac{q^m - 1}{q - 1} d$

[37]Note that the statement "(43) holds for $n = 0$" (which we just proved) is precisely the statement $\mathcal{A}(0)$ in the previous two versions of our proof.

[38]Note that the statement "(43) holds for $n = m$" (which we just assumed) is precisely the statement $\mathcal{A}(m)$ in the previous two versions of our proof.

[39]Note that this statement "(43) holds for $n = m + 1$" is precisely the statement $\mathcal{A}(m+1)$ in the previous two versions of our proof.

holds. Now, (42) (applied to $n = m$) yields

$$a_{m+1} = q \underbrace{a_m}_{= q^m a_0 + \frac{q^m - 1}{q-1} d} + d = q \left(q^m a_0 + \frac{q^m - 1}{q-1} d \right) + d$$
$$= \underbrace{qq^m}_{= q^{m+1}} a_0 + \underbrace{q \cdot \frac{q^m - 1}{q-1} d + d}_{= \left(q \cdot \frac{q^m - 1}{q-1} + 1 \right) d}$$
$$= q^{m+1} a_0 + \underbrace{\left(q \cdot \frac{q^m - 1}{q-1} + 1 \right)}_{\substack{= \frac{q(q^m - 1) + (q-1)}{q-1} = \frac{q^{m+1} - 1}{q-1} \\ (\text{since } q(q^m-1)+(q-1) = qq^m - q + q - 1 = qq^m - 1 = q^{m+1} - 1)}} d$$
$$= q^{m+1} a_0 + \frac{q^{m+1} - 1}{q-1} d.$$

So we have shown that $a_{m+1} = q^{m+1} a_0 + \dfrac{q^{m+1} - 1}{q-1} d$. In other words, (43) holds for $n = m + 1$. This completes the induction step. Hence, (43) is proven by induction. This proves Proposition 2.2. $\square$

2.2. Examples from modular arithmetic

2.2.1. Divisibility of integers

We shall soon give some more examples of inductive proofs, including some that will include slightly new tactics. These examples come from the realm of *modular arithmetic*, which is the study of congruences modulo integers. Before we come to these examples, we will introduce the definition of such congruences. But first, let us recall the definition of divisibility:

Definition 2.4. Let u and v be two integers. Then, we say that u *divides* v if and only if there exists an integer w such that $v = uw$. Instead of saying "u divides v", we can also say "v is *divisible by* u" or "v is a *multiple* of u" or "u is a *divisor* of v" or "$u \mid v$".

Thus, two integers u and v satisfy $u \mid v$ if and only if there is some $w \in \mathbb{Z}$ such that $v = uw$. For example, $1 \mid v$ holds for every integer v (since $v = 1v$), whereas $0 \mid v$ holds only for $v = 0$ (since $v = 0w$ is equivalent to $v = 0$). An integer v satisfies $2 \mid v$ if and only if v is even.

Definition 2.4 is fairly common in the modern literature (e.g., it is used in [Day16], [LeLeMe16], [Mulhol16] and [Rotman15]), but there are also some books that define these notations differently. For example, in [GrKnPa94], the notation "u divides v" is defined differently (it requires not only the existence of an integer w such that $v = uw$, but also that u is positive), whereas the notation "v is a multiple of u" is defined as it is here (i.e., it just means that there exists an integer w such that $v = uw$); thus, these two notations are not mutually interchangeable in [GrKnPa94].

Let us first prove some basic properties of divisibility:

Proposition 2.5. Let a, b and c be three integers such that $a \mid b$ and $b \mid c$. Then, $a \mid c$.

Proof of Proposition 2.5. We have $a \mid b$. In other words, there exists an integer w such that $b = aw$ (by the definition of "divides"). Consider this w, and denote it by k. Thus, k is an integer such that $b = ak$.

We have $b \mid c$. In other words, there exists an integer w such that $c = bw$ (by the definition of "divides"). Consider this w, and denote it by j. Thus, j is an integer such that $c = bj$.

Now, $c = \underbrace{b}_{=ak} j = akj$. Hence, there exists an integer w such that $c = aw$ (namely, $w = kj$). In other words, a divides c (by the definition of "divides"). In other words, $a \mid c$. This proves Proposition 2.5. □

Proposition 2.6. Let a, b and c be three integers such that $a \mid b$. Then, $ac \mid bc$.

Proof of Proposition 2.6. We have $a \mid b$. In other words, there exists an integer w such that $b = aw$ (by the definition of "divides"). Consider this w, and denote it by k. Thus, k is an integer such that $b = ak$. Hence, $\underbrace{b}_{=ak} c = akc = ack$. Thus, there exists an integer w such that $bc = acw$ (namely, $w = k$). In other words, ac divides bc (by the definition of "divides"). In other words, $ac \mid bc$. This proves Proposition 2.6. □

Proposition 2.7. Let a, b, g, x and y be integers such that $g = ax + by$. Let d be an integer such that $d \mid a$ and $d \mid b$. Then, $d \mid g$.

Proof of Proposition 2.7. We have $d \mid a$. In other words, there exists an integer w such that $a = dw$ (by the definition of "divides"). Consider this w, and denote it by p. Thus, p is an integer and satisfies $a = dp$.

Similarly, there is an integer q such that $b = dq$. Consider this q.

Now, $g = \underbrace{a}_{=dp} x + \underbrace{b}_{=dq} y = dpx + dqy = d(px + qy)$. Hence, there exists an integer w such that $g = dw$ (namely, $w = px + qy$). In other words, $d \mid g$ (by the definition of "divides"). This proves Proposition 2.7. □

It is easy to characterize divisibility in terms of fractions:

Proposition 2.8. Let a and b be two integers such that $a \neq 0$. Then, $a \mid b$ if and only if b/a is an integer.

Proof of Proposition 2.8. We first claim the following logical implication[40]:

$$(a \mid b) \implies (b/a \text{ is an integer}). \tag{45}$$

[*Proof of (45):* Assume that $a \mid b$. In other words, there exists an integer w such that $b = aw$ (by the definition of "divides"). Consider this w. Now, dividing the equality $b = aw$ by a, we obtain $b/a = w$ (since $a \neq 0$). Hence, b/a is an integer (since w is an integer). This proves the implication (45).]

Next, we claim the following logical implication:

$$(b/a \text{ is an integer}) \implies (a \mid b). \tag{46}$$

[*Proof of (46):* Assume that b/a is an integer. Let k denote this integer. Thus, $b/a = k$, so that $b = ak$. Hence, there exists an integer w such that $b = aw$ (namely, $w = k$). In other words, a divides b (by the definition of "divides"). In other words, $a \mid b$. This proves the implication (46).]

Combining the implications (45) and (46), we obtain the equivalence $(a \mid b) \iff (b/a \text{ is an integer})$. In other words, $a \mid b$ if and only if b/a is an integer. This proves Proposition 2.8. □

2.2.2. Definition of congruences

We can now define congruences:

Definition 2.9. Let a, b and n be three integers. Then, we say that *a is congruent to b modulo n* if and only if $n \mid a - b$. We shall use the notation "$a \equiv b \mod n$" for "a is congruent to b modulo n". Relations of the form "$a \equiv b \mod n$" (for integers a, b and n) are called *congruences modulo n*.

Thus, three integers a, b and n satisfy $a \equiv b \mod n$ if and only if $n \mid a - b$.

Hence, in particular:

- Any two integers a and b satisfy $a \equiv b \mod 1$. (Indeed, any two integers a and b satisfy $a - b = 1(a - b)$, thus $1 \mid a - b$, thus $a \equiv b \mod 1$.)
- Two integers a and b satisfy $a \equiv b \mod 0$ if and only if $a = b$. (Indeed, $a \equiv b \mod 0$ is equivalent to $0 \mid a - b$, which in turn is equivalent to $a - b = 0$, which in turn is equivalent to $a = b$.)
- Two integers a and b satisfy $a \equiv b \mod 2$ if and only if they have the same parity (i.e., they are either both odd or both even). This is not obvious at this point yet, but follows from Proposition 2.159 further below.

[40] A *logical implication* (or, short, *implication*) is a logical statement of the form "if $\mathcal{A}$, then $\mathcal{B}$" (where $\mathcal{A}$ and $\mathcal{B}$ are two statements).

We have

$$4 \equiv 10 \mod 3 \qquad \text{and} \qquad 5 \equiv -35 \mod 4.$$

Note that Day, in [Day16], writes "$a \equiv_n b$" instead of "$a \equiv b \mod n$". Also, other authors (particularly of older texts) write "$a \equiv b \pmod{n}$" instead of "$a \equiv b \mod n$".

Let us next introduce notations for the negations of the statements "$u \mid v$" and "$a \equiv b \mod n$":

Definition 2.10. **(a)** If u and v are two integers, then the notation "$u \nmid v$" shall mean "not $u \mid v$" (that is, "u does not divide v").

(b) If a, b and n are three integers, then the notation "$a \not\equiv b \mod n$" shall mean "not $a \equiv b \mod n$" (that is, "a is not congruent to b modulo n").

Thus, three integers a, b and n satisfy $a \not\equiv b \mod n$ if and only if $n \nmid a - b$. For example, $1 \not\equiv -1 \mod 3$, since $3 \nmid 1 - (-1)$.

2.2.3. Congruence basics

Let us now state some of the basic laws of congruences (so far, not needing induction to prove):

Proposition 2.11. Let a and n be integers. Then:

(a) We have $a \equiv 0 \mod n$ if and only if $n \mid a$.

(b) Let b be an integer. Then, $a \equiv b \mod n$ if and only if $a \equiv b \mod (-n)$.

(c) Let m and b be integers such that $m \mid n$. If $a \equiv b \mod n$, then $a \equiv b \mod m$.

Proof of Proposition 2.11. **(a)** We have the following chain of logical equivalences:

$$\begin{aligned}
&(a \equiv 0 \mod n) \\
&\iff (a \text{ is congruent to } 0 \text{ modulo } n) \\
&\qquad (\text{since "} a \equiv 0 \mod n \text{" is just a notation for "} a \text{ is congruent to } 0 \text{ modulo } n\text{"}) \\
&\iff \left(n \mid \underbrace{a - 0}_{=a}\right) \qquad (\text{by the definition of "congruent"}) \\
&\iff (n \mid a).
\end{aligned}$$

Thus, we have $a \equiv 0 \mod n$ if and only if $n \mid a$. This proves Proposition 2.11 **(a)**.

(b) Let us first assume that $a \equiv b \mod n$. Thus, a is congruent to b modulo n. In other words, $n \mid a - b$ (by the definition of "congruent"). In other words, n divides $a - b$. In other words, there exists an integer w such that $a - b = nw$ (by the definition of "divides"). Consider this w, and denote it by k. Thus, k is an integer such that $a - b = nk$.

Thus, $a - b = nk = (-n)(-k)$. Hence, there exists an integer w such that $a - b = (-n)w$ (namely, $w = -k$). In other words, $-n$ divides $a - b$ (by the definition of "divides"). In other words, $-n \mid a-b$. In other words, a is congruent to b modulo $-n$ (by the definition of "congruent"). In other words, $a \equiv b \operatorname{mod} (-n)$.

Now, forget that we assumed that $a \equiv b \operatorname{mod} n$. We thus have shown that

$$\text{if } a \equiv b \operatorname{mod} n, \text{ then } a \equiv b \operatorname{mod} (-n). \tag{47}$$

The same argument (applied to $-n$ instead of n) shows that

$$\text{if } a \equiv b \operatorname{mod} (-n), \text{ then } a \equiv b \operatorname{mod} (-(-n)).$$

Since $-(-n) = n$, this rewrites as follows:

$$\text{if } a \equiv b \operatorname{mod} (-n), \text{ then } a \equiv b \operatorname{mod} n.$$

Combining this implication with (47), we conclude that $a \equiv b \operatorname{mod} n$ if and only if $a \equiv b \operatorname{mod} (-n)$. This proves Proposition 2.11 **(b)**.

(c) Assume that $a \equiv b \operatorname{mod} n$. Thus, a is congruent to b modulo n. In other words, $n \mid a-b$ (by the definition of "congruent"). Hence, Proposition 2.5 (applied to m, n and $a-b$ instead of a, b and c) yields $m \mid a-b$ (since $m \mid n$). In other words, a is congruent to b modulo m (by the definition of "congruent"). Thus, $a \equiv b \operatorname{mod} m$. This proves Proposition 2.11 **(c)**. □

Proposition 2.12. Let n be an integer.

(a) For any integer a, we have $a \equiv a \operatorname{mod} n$.

(b) For any integers a and b satisfying $a \equiv b \operatorname{mod} n$, we have $b \equiv a \operatorname{mod} n$.

(c) For any integers a, b and c satisfying $a \equiv b \operatorname{mod} n$ and $b \equiv c \operatorname{mod} n$, we have $a \equiv c \operatorname{mod} n$.

Proof of Proposition 2.12. **(a)** Let a be an integer. Then, $a - a = 0 = n \cdot 0$. Hence, there exists an integer w such that $a - a = nw$ (namely, $w = 0$). In other words, n divides $a - a$ (by the definition of "divides"). In other words, $n \mid a - a$. In other words, a is congruent to a modulo n (by the definition of "congruent"). In other words, $a \equiv a \operatorname{mod} n$. This proves Proposition 2.12 **(a)**.

(b) Let a and b be two integers satisfying $a \equiv b \operatorname{mod} n$. Thus, a is congruent to b modulo n (since $a \equiv b \operatorname{mod} n$). In other words, $n \mid a - b$ (by the definition of "congruent"). In other words, n divides $a - b$. In other words, there exists an integer w such that $a - b = nw$ (by the definition of "divides"). Consider this w, and denote it by q. Thus, q is an integer such that $a - b = nq$. Now, $b - a = -\underbrace{(a-b)}_{=nq} = -nq = n(-q)$. Hence, there exists an integer w such that $b - a = nw$ (namely, $w = -q$). In other words, n divides $b - a$ (by the definition of "divides"). In other words, $n \mid b - a$. In other words, b is congruent to a modulo n (by the definition of "congruent"). In other words, $b \equiv a \operatorname{mod} n$. This proves Proposition 2.12 **(b)**.

(c) Let a, b and c be three integers satisfying $a \equiv b \operatorname{mod} n$ and $b \equiv c \operatorname{mod} n$.

Just as in the above proof of Proposition 2.12 **(b)**, we can use the assumption $a \equiv b \mod n$ to construct an integer q such that $a - b = nq$. Similarly, we can use the assumption $b \equiv c \mod n$ to construct an integer r such that $b - c = nr$. Consider these q and r.

Now,

$$a - c = \underbrace{(a - b)}_{=nq} + \underbrace{(b - c)}_{=nr} = nq + nr = n(q + r).$$

Hence, there exists an integer w such that $a - c = nw$ (namely, $w = q + r$). In other words, n divides $a - c$ (by the definition of "divides"). In other words, $n \mid a - c$. In other words, a is congruent to c modulo n (by the definition of "congruent"). In other words, $a \equiv c \mod n$. This proves Proposition 2.12 **(c)**. □

Simple as they are, the three parts of Proposition 2.12 have names: Proposition 2.12 **(a)** is called the *reflexivity of congruence (modulo n)*; Proposition 2.12 **(b)** is called the *symmetry of congruence (modulo n)*; Proposition 2.12 **(c)** is called the *transitivity of congruence (modulo n)*.

Proposition 2.12 **(b)** allows the following definition:

Definition 2.13. Let n, a and b be three integers. Then, we say that *a and b are congruent modulo n* if and only if $a \equiv b \mod n$. Proposition 2.12 **(b)** shows that a and b actually play equal roles in this relation (i.e., the statement "a and b are congruent modulo n" is equivalent to "b and a are congruent modulo n").

Proposition 2.14. Let n be an integer. Then, $n \equiv 0 \mod n$.

Proof of Proposition 2.14. We have $n = n \cdot 1$. Thus, there exists an integer w such that $n = nw$ (namely, $w = 1$). Therefore, $n \mid n$ (by the definition of "divides"). Proposition 2.11 **(a)** (applied to $a = n$) shows that we have $n \equiv 0 \mod n$ if and only if $n \mid n$. Hence, we have $n \equiv 0 \mod n$ (since $n \mid n$). This proves Proposition 2.14. □

2.2.4. Chains of congruences

Proposition 2.12 shows that congruences (modulo n) behave like equalities – in that we can turn them around (since Proposition 2.12 **(b)** says that $a \equiv b \mod n$ implies $b \equiv a \mod n$) and we can chain them together (by Proposition 2.12 **(c)**) and in that every integer is congruent to itself (by Proposition 2.12 **(a)**). This leads to the following notation:

Definition 2.15. If $a_1, a_2, \ldots, a_k$ and n are integers, then the statement "$a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n$" shall mean that

$$(a_i \equiv a_{i+1} \mod n \text{ holds for each } i \in \{1, 2, \ldots, k-1\}).$$

Such a statement is called a *chain of congruences modulo n* (or, less precisely, a *chain of congruences*). We shall refer to the integers $a_1, a_2, \ldots, a_k$ (but not n) as the *members* of this chain.

For example, the chain $a \equiv b \equiv c \equiv d \mod n$ (for five integers a, b, c, d, n) means that $a \equiv b \mod n$ and $b \equiv c \mod n$ and $c \equiv d \mod n$.

The usefulness of such chains lies in the following fact:

Proposition 2.16. Let $a_1, a_2, \ldots, a_k$ and n be integers such that $a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n$. Let u and v be two elements of $\{1, 2, \ldots, k\}$. Then,

$$a_u \equiv a_v \mod n.$$

In other words, any two members of a chain of congruences modulo n are congruent to each other modulo n. Thus, chains of congruences are like chains of equalities: From any chain of congruences modulo n with k members, you can extract k^2 congruences modulo n by picking any two members of the chain.

Example 2.17. Proposition 2.16 shows (among other things) that if a, b, c, d, e, n are integers such that $a \equiv b \equiv c \equiv d \equiv e \mod n$, then $a \equiv d \mod n$ and $b \equiv d \mod n$ and $e \equiv b \mod n$ (and various other congruences).

Unsurprisingly, Proposition 2.16 can be proven by induction, although not in an immediately obvious manner: We cannot directly prove it by induction on n, on k, on u or on v. Instead, we will first introduce an auxiliary statement (the statement (49) in the following proof) which will be tailored to an inductive proof. This is a commonly used tactic, and particularly helpful to us now as we only have the most basic form of the principle of induction available. (Soon, we will see more versions of that principle, which will obviate the need for some of the tailoring.)

Proof of Proposition 2.16. By assumption, we have $a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n$. In other words,

$$(a_i \equiv a_{i+1} \mod n \text{ holds for each } i \in \{1, 2, \ldots, k-1\}) \tag{48}$$

(since this is what "$a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n$" means).

Fix $p \in \{1, 2, \ldots, k\}$. For each $i \in \mathbb{N}$, we let $\mathcal{A}(i)$ be the statement

$$\left(\text{if } p + i \in \{1, 2, \ldots, k\}, \text{ then } a_p \equiv a_{p+i} \mod n\right). \tag{49}$$

We shall prove that this statement $\mathcal{A}(i)$ holds for each $i \in \mathbb{N}$.

In fact, let us prove this by induction on i: [41]

Induction base: The statement $\mathcal{A}(0)$ holds[42]. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that $\mathcal{A}(m)$ holds. We must show that $\mathcal{A}(m+1)$ holds.

[41] Thus, the letter "i" plays the role of the "n" in Theorem 2.1 (since we are already using "n" for a different thing).

[42] *Proof.* Proposition 2.12 **(a)** (applied to $a = a_p$) yields $a_p \equiv a_p \mod n$. In view of $p = p + 0$, this rewrites as $a_p \equiv a_{p+0} \mod n$. Hence, $\left(\text{if } p + 0 \in \{1, 2, \ldots, k\}, \text{ then } a_p \equiv a_{p+0} \mod n\right)$. But this is precisely the statement $\mathcal{A}(0)$. Hence, the statement $\mathcal{A}(0)$ holds.

We have assumed that $\mathcal{A}(m)$ holds. In other words,

$$\left(\text{if } p+m \in \{1,2,\ldots,k\}, \text{ then } a_p \equiv a_{p+m} \operatorname{mod} n\right). \tag{50}$$

Next, let us assume that $p+(m+1) \in \{1,2,\ldots,k\}$. Thus, $p+(m+1) \leq k$, so that $p+m+1 = p+(m+1) \leq k$ and therefore $p+m \leq k-1$. Also, $p \in \{1,2,\ldots,k\}$, so that $p \geq 1$ and thus $\underbrace{p}_{\geq 1} + \underbrace{m}_{\geq 0} \geq 1+0 = 1$. Combining this with $p+m \leq k-1$, we obtain $p+m \in \{1,2,\ldots,k-1\} \subseteq \{1,2,\ldots,k\}$. Hence, (50) shows that $a_p \equiv a_{p+m} \operatorname{mod} n$. But (48) (applied to $p+m$ instead of i) yields $a_{p+m} \equiv a_{(p+m)+1} \operatorname{mod} n$ (since $p+m \in \{1,2,\ldots,k-1\}$).

So we know that $a_p \equiv a_{p+m} \operatorname{mod} n$ and $a_{p+m} \equiv a_{(p+m)+1} \operatorname{mod} n$. Hence, Proposition 2.12 **(c)** (applied to $a = a_p$, $b = a_{p+m}$ and $c = a_{(p+m)+1}$) yields $a_p \equiv a_{(p+m)+1} \operatorname{mod} n$. Since $(p+m)+1 = p+(m+1)$, this rewrites as $a_p \equiv a_{p+(m+1)} \operatorname{mod} n$.

Now, forget that we assumed that $p+(m+1) \in \{1,2,\ldots,k\}$. We thus have shown that

$$\left(\text{if } p+(m+1) \in \{1,2,\ldots,k\}, \text{ then } a_p \equiv a_{p+(m+1)} \operatorname{mod} n\right).$$

But this is precisely the statement $\mathcal{A}(m+1)$. Thus, $\mathcal{A}(m+1)$ holds.

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{N}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. This completes the induction step.

Thus, we have completed both the induction base and the induction step. Hence, by induction, we conclude that $\mathcal{A}(i)$ holds for each $i \in \mathbb{N}$. In other words, (49) holds for each $i \in \mathbb{N}$.

We are not done yet, since our goal is to prove Proposition 2.16, not merely to prove $\mathcal{A}(i)$. But this is now easy.

First, let us forget that we fixed p. Thus, we have shown that (49) holds for each $p \in \{1,2,\ldots,k\}$ and $i \in \mathbb{N}$.

But we have either $u \leq v$ or $u > v$. In other words, we are in one of the following two cases:

Case 1: We have $u \leq v$.

Case 2: We have $u > v$.

Let us first consider Case 1. In this case, we have $u \leq v$. Thus, $v-u \geq 0$, so that $v-u \in \mathbb{N}$. But recall that (49) holds for each $p \in \{1,2,\ldots,k\}$ and $i \in \mathbb{N}$. Applying this to $p = u$ and $i = v-u$, we conclude that (49) holds for $p = u$ and $i = v-u$ (since $u \in \{1,2,\ldots,k\}$ and $v-u \in \mathbb{N}$). In other words,

$$\left(\text{if } u+(v-u) \in \{1,2,\ldots,k\}, \text{ then } a_u \equiv a_{u+(v-u)} \operatorname{mod} n\right).$$

Since $u+(v-u) = v$, this rewrites as

$$\left(\text{if } v \in \{1,2,\ldots,k\}, \text{ then } a_u \equiv a_v \operatorname{mod} n\right).$$

Since $v \in \{1,2,\ldots,k\}$ holds (by assumption), we conclude that $a_u \equiv a_v \operatorname{mod} n$. Thus, Proposition 2.16 is proven in Case 1.

Let us now consider Case 2. In this case, we have $u > v$. Thus, $u - v > 0$, so that $u - v \in \mathbb{N}$. But recall that (49) holds for each $p \in \{1, 2, \ldots, k\}$ and $i \in \mathbb{N}$. Applying this to $p = v$ and $i = u - v$, we conclude that (49) holds for $p = v$ and $i = u - v$ (since $v \in \{1, 2, \ldots, k\}$ and $u - v \in \mathbb{N}$). In other words,

$$\left(\text{if } v + (u - v) \in \{1, 2, \ldots, k\}, \text{ then } a_v \equiv a_{v+(u-v)} \bmod n\right).$$

Since $v + (u - v) = u$, this rewrites as

$$(\text{if } u \in \{1, 2, \ldots, k\}, \text{ then } a_v \equiv a_u \bmod n).$$

Since $u \in \{1, 2, \ldots, k\}$ holds (by assumption), we conclude that $a_v \equiv a_u \bmod n$. Therefore, Proposition 2.12 **(b)** (applied to $a = a_v$ and $b = a_u$) yields that $a_u \equiv a_v \bmod n$. Thus, Proposition 2.16 is proven in Case 2.

Hence, Proposition 2.16 is proven in both Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Proposition 2.16 always holds. □

2.2.5. Chains of inequalities (a digression)

Much of the above proof of Proposition 2.16 was unremarkable and straightforward reasoning – but this proof is nevertheless fundamental and important. More or less the same argument can be used to show the following fact about chains of inequalities:

Proposition 2.18. Let $a_1, a_2, \ldots, a_k$ be integers such that $a_1 \leq a_2 \leq \cdots \leq a_k$. (Recall that the statement "$a_1 \leq a_2 \leq \cdots \leq a_k$" means that ($a_i \leq a_{i+1}$ holds for each $i \in \{1, 2, \ldots, k-1\}$).) Let u and v be two elements of $\{1, 2, \ldots, k\}$ such that $u \leq v$. Then,

$$a_u \leq a_v.$$

Proposition 2.18 is similar to Proposition 2.16, with the congruences replaced by inequalities; but note that the condition "$u \leq v$" is now required. Make sure you understand where you need this condition when adapting the proof of Proposition 2.16 to Proposition 2.18!

For future use, let us prove a corollary of Proposition 2.18 which essentially observes that the inequality sign in $a_u \leq a_v$ can be made strict if there is any strict inequality sign between a_u and a_v in the chain $a_1 \leq a_2 \leq \cdots \leq a_k$:

Corollary 2.19. Let $a_1, a_2, \ldots, a_k$ be integers such that $a_1 \leq a_2 \leq \cdots \leq a_k$. Let u and v be two elements of $\{1, 2, \ldots, k\}$ such that $u \leq v$. Let $p \in \{u, u+1, \ldots, v-1\}$ be such that $a_p < a_{p+1}$. Then,

$$a_u < a_v.$$

Proof of Corollary 2.19. From $u \in \{1, 2, \ldots, k\}$, we obtain $u \geq 1$. From $v \in \{1, 2, \ldots, k\}$, we obtain $v \leq k$. From $p \in \{u, u+1, \ldots, v-1\}$, we obtain $p \geq u$ and $p \leq v-1$. From $p \leq v-1$, we obtain $p+1 \leq v \leq k$. Combining this with $p+1 \geq p \geq u \geq 1$, we obtain $p+1 \in \{1, 2, \ldots, k\}$ (since $p+1$ is an integer). Combining $p \leq v-1 \leq v \leq k$ with $p \geq u \geq 1$, we obtain $p \in \{1, 2, \ldots, k\}$ (since p is an integer). We thus know that both p and $p+1$ are elements of $\{1, 2, \ldots, k\}$.

We have $p \geq u$, thus $u \leq p$. Hence, Proposition 2.18 (applied to p instead of v) yields $a_u \leq a_p$. Combining this with $a_p < a_{p+1}$, we find $a_u < a_{p+1}$.

We have $p+1 \leq v$. Hence, Proposition 2.18 (applied to $p+1$ instead of u) yields $a_{p+1} \leq a_v$. Combining $a_u < a_{p+1}$ with $a_{p+1} \leq a_v$, we obtain $a_u < a_v$. This proves Corollary 2.19. □

In particular, we see that the inequality sign in $a_u \leq a_v$ is strict when $u < v$ holds and **all** inequality signs in the chain $a_1 \leq a_2 \leq \cdots \leq a_k$ are strict:

Corollary 2.20. Let $a_1, a_2, \ldots, a_k$ be integers such that $a_1 < a_2 < \cdots < a_k$. (Recall that the statement "$a_1 < a_2 < \cdots < a_k$" means that ($a_i < a_{i+1}$ holds for each $i \in \{1, 2, \ldots, k-1\}$).) Let u and v be two elements of $\{1, 2, \ldots, k\}$ such that $u < v$. Then,

$$a_u < a_v.$$

Proof of Corollary 2.20. From $u < v$, we obtain $u \leq v-1$ (since u and v are integers). Combining this with $u \geq u$, we conclude that $u \in \{u, u+1, \ldots, v-1\}$. Also, from $a_1 < a_2 < \cdots < a_k$, we obtain $a_1 \leq a_2 \leq \cdots \leq a_k$.

We have $u \leq v-1$, thus $u+1 \leq v \leq k$ (since $v \in \{1, 2, \ldots, k\}$), so that $u \leq k-1$. Combining this with $u \geq 1$ (which is a consequence of $u \in \{1, 2, \ldots, k\}$), we find $u \in \{1, 2, \ldots, k-1\}$. Hence, from $a_1 < a_2 < \cdots < a_k$, we obtain $a_u < a_{u+1}$. Hence, Corollary 2.19 (applied to $p = u$) yields $a_u < a_v$ (since $u \leq v$ (because $u < v$)). This proves Corollary 2.20. □

2.2.6. Addition, subtraction and multiplication of congruences

Let us now return to the topic of congruences.

Chains of congruences can include equality signs. For example, if a, b, c, d, n are integers, then "$a \equiv b = c \equiv d \mod n$" means that $a \equiv b \mod n$ and $b = c$ and $c \equiv d \mod n$. Such a chain is still a chain of congruences, because $b = c$ implies $b \equiv c \mod n$ (by Proposition 2.12 **(a)**).

Let us continue with basic properties of congruences:

Proposition 2.21. Let a, b, c, d and n be integers such that $a \equiv b \mod n$ and $c \equiv d \mod n$. Then:

(a) We have $a + c \equiv b + d \mod n$.

(b) We have $a - c \equiv b - d \mod n$.

(c) We have $ac \equiv bd \mod n$.

Note that Proposition 2.21 does **not** claim that $a/c \equiv b/d \mod n$. Indeed, this would not be true in general. One reason for this is that a/c and b/d aren't always integers. But even when they are, they may not satisfy $a/c \equiv b/d \mod n$. For example, $6 \equiv 4 \mod 2$ and $2 \equiv 2 \mod 2$, but $6/2 \not\equiv 4/2 \mod 2$. Likewise, Proposition 2.21 does **not** claim that $a^c \equiv b^d \mod n$ even when a, b, c, d are nonnegative; that too would not be true. But we will soon see that a weaker statement (Proposition 2.22) holds. First, let us prove Proposition 2.21:

Proof of Proposition 2.21. From $a \equiv b \mod n$, we conclude that a is congruent to b modulo n. In other words, $n \mid a - b$ (by the definition of "congruent"). In other words, n divides $a - b$. In other words, there exists an integer w such that $a - b = nw$ (by the definition of "divides"). Consider this w, and denote it by q. Thus, q is an integer such that $a - b = nq$.

Similarly, from $c \equiv d \mod n$, we can construct an integer r such that $c - d = nr$. Consider this r.

(a) We have

$$(a+c) - (b+d) = \underbrace{(a-b)}_{=nq} + \underbrace{(c-d)}_{=nr} = nq + nr = n(q+r).$$

Hence, there exists an integer w such that $(a+c) - (b+d) = nw$ (namely, $w = q + r$). In other words, n divides $(a+c) - (b+d)$ (by the definition of "divides"). In other words, $n \mid (a+c) - (b+d)$. In other words, $a + c \equiv b + d \mod n$ (by the definition of "congruent"). This proves Proposition 2.21 **(a)**.

(b) We have

$$(a-c) - (b-d) = \underbrace{(a-b)}_{=nq} - \underbrace{(c-d)}_{=nr} = nq - nr = n(q-r).$$

Hence, there exists an integer w such that $(a-c) - (b-d) = nw$ (namely, $w = q - r$). In other words, n divides $(a-c) - (b-d)$ (by the definition of "divides"). In other words, $n \mid (a-c) - (b-d)$. In other words, $a - c \equiv b - d \mod n$ (by the definition of "congruent"). This proves Proposition 2.21 **(b)**.

(c) We have $ac - ad = a\underbrace{(c-d)}_{=nr} = anr = n(ar)$. Hence, there exists an integer w such that $ac - ad = nw$ (namely, $w = ar$). In other words, n divides $ac - ad$ (by the definition of "divides"). In other words, $n \mid ac - ad$. In other words, $ac \equiv ad \mod n$ (by the definition of "congruent").

We have $ad - bd = \underbrace{(a-b)}_{=nq} d = nqd = n(qd)$. Hence, there exists an integer w such that $ad - bd = nw$ (namely, $w = qd$). In other words, n divides $ad - bd$ (by the definition of "divides"). In other words, $n \mid ad - bd$. In other words, $ad \equiv bd \mod n$ (by the definition of "congruent").

Now, we know that $ac \equiv ad \mod n$ and $ad \equiv bd \mod n$. Hence, Proposition 2.12 **(c)** (applied to ac, ad and bd instead of a, b and c) shows that $ac \equiv bd \mod n$. This proves Proposition 2.21 **(c)**. □

Proposition 2.21 shows yet another aspect in which congruences (modulo n) behave like equalities: They can be added, subtracted and multiplied, in the following sense:

- We can add two congruences modulo n (in the sense of adding each side of one congruence to the corresponding side of the other); this yields a new congruence modulo n (because of Proposition 2.21 **(a)**).
- We can subtract two congruences modulo n; this yields a new congruence modulo n (because of Proposition 2.21 **(b)**).
- We can multiply two congruences modulo n; this yields a new congruence modulo n (because of Proposition 2.21 **(c)**).

2.2.7. Substitutivity for congruences

Combined with Proposition 2.12, these observations lead to a further feature of congruences, which is even more important: the principle of *substitutivity for congruences*. We are not going to state it fully formally (as it is a meta-mathematical principle), but merely explain what it means.

Recall that the *principle of substitutivity for equalities* says the following:

> *Principle of substitutivity for equalities:* If two objects[43] x and x' are equal, and if we have any expression A that involves the object x, then we can replace this x (or, more precisely, any arbitrary appearance of x in A) in A by x'; the value of the resulting expression A' will be equal to the value of A.

Here are two examples of how this principle can be used:

- If a, b, c, d, e, c' are numbers such that $c = c'$, then the principle of substitutivity for equalities says that we can replace c by c' in the expression $a(b-(c+d)e)$, and the value of the resulting expression $a(b-(c'+d)e)$ will be equal to the value of $a(b-(c+d)e)$; that is, we have
$$a(b-(c+d)e) = a\left(b-\left(c'+d\right)e\right). \tag{51}$$
- If a, b, c, a' are numbers such that $a = a'$, then
$$(a-b)(a+b) = \left(a'-b\right)(a+b), \tag{52}$$
because the principle of substitutivity allows us to replace the first a appearing in the expression $(a-b)(a+b)$ by an a'. (We can also replace the second a by a', of course.)

[43]"Objects" can be numbers, sets, tuples or any other mathematical objects.

More generally, we can make several such replacements at the same time.

The principle of substitutivity for equalities is one of the headstones of mathematical logic; it is the essence of what it means for two objects to be equal.

The *principle of substitutivity for congruences* is similar, but far less fundamental; it says the following:

> *Principle of substitutivity for congruences:* Fix an integer n. If two numbers x and x' are congruent to each other modulo n (that is, $x \equiv x' \mod n$), and if we have any expression A that involves only integers, addition, subtraction and multiplication, and involves the object x, then we can replace this x (or, more precisely, any arbitrary appearance of x in A) in A by x'; the value of the resulting expression A' will be congruent to the value of A modulo n.

Note that this principle is less general than the principle of substitutivity for equalities, because it only applies to expressions that are built from integers and certain operations (note that division is not one of these operations). But it still lets us prove analogues of our above examples (51) and (52):

- If n is any integer, and if a, b, c, d, e, c' are integers such that $c \equiv c' \mod n$, then the principle of substitutivity for congruences says that we can replace c by c' in the expression $a (b - (c + d) e)$, and the value of the resulting expression $a (b - (c' + d) e)$ will be congruent to the value of $a (b - (c + d) e)$ modulo n; that is, we have

$$a (b - (c + d) e) \equiv a \left(b - \left(c' + d\right) e\right) \mod n. \tag{53}$$

- If n is any integer, and if a, b, c, a' are integers such that $a \equiv a' \mod n$, then

$$(a - b) (a + b) \equiv \left(a' - b\right) (a + b) \mod n, \tag{54}$$

  because the principle of substitutivity allows us to replace the first a appearing in the expression $(a - b) (a + b)$ by an a'. (We can also replace the second a by a', of course.)

We shall not prove the principle of substitutivity for congruences, since we have not formalized it (after all, we have not defined what an "expression" is). But we shall prove the specific congruences (53) and (54) using Proposition 2.21 and Proposition 2.12; the way in which we prove these congruences is symptomatic: Every congruence obtained from the principle of substitutivity for congruences can be proven in a manner like these. Thus, we hope that the proofs of (53) and (54) given below serve as templates which can easily be adapted to any other situation in which an application of the principle of substitutivity for congruences needs to be justified.

Proof of (53). Let n be any integer, and let a, b, c, d, e, c' be integers such that $c \equiv c' \operatorname{mod} n$.

Adding the congruence $c \equiv c' \operatorname{mod} n$ with the congruence $d \equiv d \operatorname{mod} n$ (which follows from Proposition 2.12 **(a)**), we obtain $c + d \equiv c' + d \operatorname{mod} n$. Multiplying this congruence with the congruence $e \equiv e \operatorname{mod} n$ (which follows from Proposition 2.12 **(a)**), we obtain $(c+d) e \equiv (c'+d) e \operatorname{mod} n$. Subtracting this congruence from the congruence $b \equiv b \operatorname{mod} n$ (which, again, follows from Proposition 2.12 **(a)**), we obtain $b - (c+d) e \equiv b - (c'+d) e \operatorname{mod} n$. Multiplying the congruence $a \equiv a \operatorname{mod} n$ (which follows from Proposition 2.12 **(a)**) with this congruence, we obtain $a (b - (c+d) e) \equiv a (b - (c'+d) e) \operatorname{mod} n$. This proves (53). □

Proof of (54). Let n be any integer, and let a, b, c, a' be integers such that $a \equiv a' \operatorname{mod} n$.

Subtracting the congruence $b \equiv b \operatorname{mod} n$ (which follows from Proposition 2.12 **(a)**) from the congruence $a \equiv a' \operatorname{mod} n$, we obtain $a - b \equiv a' - b \operatorname{mod} n$. Multiplying this congruence with the congruence $a + b \equiv a + b \operatorname{mod} n$ (which follows from Proposition 2.12 **(a)**), we obtain $(a-b)(a+b) \equiv (a'-b)(a+b) \operatorname{mod} n$. This proves (54). □

As we said, these two proofs are exemplary: Any congruence obtained from the principle of substitutivity for congruences can be proven in such a way (starting with the congruence $x \equiv x' \operatorname{mod} n$, and then "wrapping" it up in the expression A by repeatedly adding, multiplying and subtracting congruences that follow from Proposition 2.12 **(a)**).

When we apply the principle of substitutivity for congruences, we shall use underbraces to point out which integers we are replacing. For example, when deriving (53) from this principle, we shall write

$$a \left(b - \left(\underbrace{c}_{\equiv c' \operatorname{mod} n} + d \right) e \right) \equiv a \left(b - \left(c' + d \right) e \right) \operatorname{mod} n,$$

in order to stress that we are replacing c by c'. Likewise, when deriving (54) from this principle, we shall write

$$\left(\underbrace{a}_{\equiv a' \operatorname{mod} n} - b \right) (a + b) \equiv \left(a' - b \right) (a + b) \operatorname{mod} n,$$

in order to stress that we are replacing the first a (but not the second a) by a'.

The principle of substitutivity for congruences allows us to replace a **single** integer x appearing in an expression by another integer x' that is congruent to x modulo n. Applying this principle many times, we thus conclude that we can also replace **several** integers at the same time (because we can get to the same result by performing these replacements one at a time, and Proposition 2.16 shows that the final result will be congruent to the original result).

For example, if seven integers a, a', b, b', c, c', n satisfy $a \equiv a' \mod n$ and $b \equiv b' \mod n$ and $c \equiv c' \mod n$, then

$$bc + ca + ab \equiv b'c' + c'a' + a'b' \mod n, \tag{55}$$

because we can replace all the six integers b, c, c, a, a, b in the expression $bc + ca + ab$ (listed in the order of their appearance in this expression) by b', c', c', a', a', b', respectively. If we want to derive this from the principle of substitutivity for congruences, we must perform the replacements one at a time, e.g., as follows:

$$\begin{aligned}
\underbrace{b}_{\equiv b' \mod n} c + ca + ab &\equiv b' \underbrace{c}_{\equiv c' \mod n} + ca + ab \equiv b'c' + \underbrace{c}_{\equiv c' \mod n} a + ab \\
&\equiv b'c' + c' \underbrace{a}_{\equiv a' \mod n} + ab \equiv b'c' + c'a' + \underbrace{a}_{\equiv a' \mod n} b \\
&\equiv b'c' + c'a' + a' \underbrace{b}_{\equiv b' \mod n} \equiv b'c' + c'a' + a'b' \mod n.
\end{aligned}$$

Of course, we shall always just show the replacements as a single step:

$$\underbrace{b}_{\equiv b' \mod n} \underbrace{c}_{\equiv c' \mod n} + \underbrace{c}_{\equiv c' \mod n} \underbrace{a}_{\equiv a' \mod n} + \underbrace{a}_{\equiv a' \mod n} \underbrace{b}_{\equiv b' \mod n} \equiv b'c' + c'a' + a'b' \mod n.$$

2.2.8. Taking congruences to the k-th power

We have seen that congruences (like equalities) can be added, subtracted and multiplied (but, unlike equalities, they cannot be divided). One other thing we can do with congruences is taking powers of them, as long as the exponent is a nonnegative integer. This relies on the following fact:

Proposition 2.22. Let a, b and n be three integers such that $a \equiv b \mod n$. Then, $a^k \equiv b^k \mod n$ for each $k \in \mathbb{N}$.

The following proof of Proposition 2.22 is an example of a straightforward inductive proof; the only thing to keep in mind is that it uses induction on k, not induction on n as some of our previous proofs did.

Proof of Proposition 2.22. We claim that

$$a^k \equiv b^k \mod n \qquad \text{for each } k \in \mathbb{N}. \tag{56}$$

We shall prove (56) by induction on k:

Induction base: We have $1 \equiv 1 \mod n$ (by Proposition 2.12 **(a)**). In view of $a^0 = 1$ and $b^0 = 1$, this rewrites as $a^0 \equiv b^0 \mod n$. In other words, (56) holds for $k = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (56) holds for $k = m$. We must show that (56) holds for $k = m + 1$.

We have assumed that (56) holds for $k = m$. In other words, we have $a^m \equiv b^m \mod n$. Now,

$$a^{m+1} = \underbrace{a^m}_{\equiv b^m \mod n} \underbrace{a}_{\equiv b \mod n} \equiv b^m b = b^{m+1} \mod n.$$

[44] In other words, (56) holds for $k = m + 1$. This completes the induction step. Hence, (56) is proven by induction. This proves Proposition 2.22. □

2.3. A few recursively defined sequences

2.3.1. $a_n = a_{n-1}^q + r$

We next proceed to give some more examples of proofs by induction.

Example 2.23. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of integers defined recursively by

$$a_0 = 0, \qquad \text{and}$$
$$a_n = a_{n-1}^2 + 1 \qquad \text{for each } n \geq 1.$$

("Defined recursively" means that we aren't defining each entry a_n of our sequence by an explicit formula, but rather defining a_n in terms of the previous entries $a_0, a_1, \ldots, a_{n-1}$. Thus, in order to compute some entry a_n of our sequence, we need to compute all the previous entries $a_0, a_1, \ldots, a_{n-1}$. This means that if we want to compute a_n, we should first compute a_0, then compute a_1 (using our value of a_0), then compute a_2 (using our values of a_0 and a_1), and so on, until we reach a_n. For example, in order to compute a_6, we proceed as follows:

$$\begin{aligned}
a_0 &= 0; \\
a_1 &= a_0^2 + 1 = 0^2 + 1 = 1; \\
a_2 &= a_1^2 + 1 = 1^2 + 1 = 2; \\
a_3 &= a_2^2 + 1 = 2^2 + 1 = 5; \\
a_4 &= a_3^2 + 1 = 5^2 + 1 = 26; \\
a_5 &= a_4^2 + 1 = 26^2 + 1 = 677; \\
a_6 &= a_5^2 + 1 = 677^2 + 1 = 458\,330.
\end{aligned}$$

And similarly we can compute a_n for any $n \in \mathbb{N}$.)

This sequence $(a_0, a_1, a_2, \ldots)$ is not unknown: It is the sequence A003095 in the Online Encyclopedia of Integer Sequences.

[44] This computation relied on the principle of substitutivity for congruences. Here is how to rewrite this argument in a more explicit way (without using this principle): We have $a^m \equiv b^m \mod n$ and $a \equiv b \mod n$. Hence, Proposition 2.21 **(c)** (applied to a^m, b^m, a and b instead of a, b, c and d) yields $a^m a \equiv b^m b \mod n$. This rewrites as $a^{m+1} \equiv b^{m+1} \mod n$ (since $a^{m+1} = a^m a$ and $b^{m+1} = b^m b$).

A look at the first few entries of the sequence makes us realize that both a_2 and a_3 divide a_6, just as the integers 2 and 3 themselves divide 6. This suggests that we might have $a_u \mid a_v$ whenever u and v are two nonnegative integers satisfying $u \mid v$. We shall soon prove this observation (which was found by Michael Somos in 2013) in greater generality.

Theorem 2.24. Fix some $q \in \mathbb{N}$ and $r \in \mathbb{Z}$. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of integers defined recursively by

$$a_0 = 0, \qquad \text{and}$$
$$a_n = a_{n-1}^q + r \qquad \text{for each } n \geq 1.$$

(Note that if $q = 2$ and $r = 1$, then this sequence $(a_0, a_1, a_2, \ldots)$ is precisely the sequence $(a_0, a_1, a_2, \ldots)$ from Example 2.23. If $q = 3$ and $r = 1$, then our sequence $(a_0, a_1, a_2, \ldots)$ is the sequence A135361 in the Online Encyclopedia of Integer Sequences. If $q = 0$, then our sequence $(a_0, a_1, a_2, \ldots)$ is $(0, r+1, r+1, r+1, \ldots)$. If $q = 1$, then our sequence $(a_0, a_1, a_2, \ldots)$ is $(0, r, 2r, 3r, 4r, \ldots)$, as can be easily proven by induction.)

(a) For any $k \in \mathbb{N}$ and $n \in \mathbb{N}$, we have $a_{k+n} \equiv a_k \operatorname{mod} a_n$.

(b) For any $n \in \mathbb{N}$ and $w \in \mathbb{N}$, we have $a_n \mid a_{nw}$.

(c) If u and v are two nonnegative integers satisfying $u \mid v$, then $a_u \mid a_v$.

Proof of Theorem 2.24. **(a)** Let $n \in \mathbb{N}$. We claim that

$$a_{k+n} \equiv a_k \operatorname{mod} a_n \qquad \text{for every } k \in \mathbb{N}. \tag{57}$$

We shall prove (57) by induction on k:

Induction base: Proposition 2.14 (applied to a_n instead of n) yields $a_n \equiv 0 \operatorname{mod} a_n$. This rewrites as $a_n \equiv a_0 \operatorname{mod} a_n$ (since $a_0 = 0$). In other words, $a_{0+n} \equiv a_0 \operatorname{mod} a_n$ (since $0 + n = n$). In other words, (57) holds for $k = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (57) holds for $k = m$. We must prove that (57) holds for $k = m + 1$.

We have assumed that (57) holds for $k = m$. In other words, we have

$$a_{m+n} \equiv a_m \operatorname{mod} a_n.$$

Hence, Proposition 2.22 (applied to a_{m+n}, a_m, a_n and q instead of a, b, n and k) shows that $a_{m+n}^q \equiv a_m^q \operatorname{mod} a_n$. Hence, $a_{m+n}^q + r \equiv a_m^q + r \operatorname{mod} a_n$. (Indeed, this follows by adding the congruence $a_{m+n}^q \equiv a_m^q \operatorname{mod} a_n$ to the congruence $r \equiv r \operatorname{mod} a_n$; the latter congruence is a consequence of Proposition 2.12 **(a)**.)

Now, $(m+1) + n = (m+n) + 1 \geq 1$. Hence, the recursive definition of the sequence $(a_0, a_1, a_2, \ldots)$ yields

$$a_{(m+1)+n} = a_{((m+1)+n)-1}^q + r = a_{m+n}^q + r$$

(since $((m+1)+n)-1 = m+n$). Also, $m+1 \geq 1$. Hence, the recursive definition of the sequence $(a_0, a_1, a_2, \ldots)$ yields

$$a_{m+1} = a_{(m+1)-1}^q + r = a_m^q + r.$$

The congruence $a_{m+n}^q + r \equiv a_m^q + r \operatorname{mod} a_n$ rewrites as $a_{(m+1)+n} \equiv a_{m+1} \operatorname{mod} a_n$ (since $a_{(m+1)+n} = a_{m+n}^q + r$ and $a_{m+1} = a_m^q + r$). In other words, (57) holds for $k = m+1$. This completes the induction step. Thus, (57) is proven by induction.

Therefore, Theorem 2.24 **(a)** is proven.

(b) Let $n \in \mathbb{N}$. We claim that

$$a_n \mid a_{nw} \qquad \text{for every } w \in \mathbb{N}. \tag{58}$$

We shall prove (58) by induction on w:

Induction base: We have $a_{n \cdot 0} = a_0 = 0$. But $a_n \mid 0$ (since $0 = 0 a_n$). This rewrites as $a_n \mid a_{n \cdot 0}$ (since $a_{n \cdot 0} = 0$). In other words, (58) holds for $w = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (58) holds for $w = m$. We must prove that (58) holds for $w = m+1$.

We have assumed that (58) holds for $w = m$. In other words, we have $a_n \mid a_{nm}$.

Proposition 2.11 **(a)** (applied to a_{nm} and a_n instead of a and n) shows that we have $a_{nm} \equiv 0 \operatorname{mod} a_n$ if and only if $a_n \mid a_{nm}$. Hence, we have $a_{nm} \equiv 0 \operatorname{mod} a_n$ (since $a_n \mid a_{nm}$).

Theorem 2.24 **(a)** (applied to $k = nm$) yields $a_{nm+n} \equiv a_{nm} \operatorname{mod} a_n$. Thus, $a_{nm+n} \equiv a_{nm} \equiv 0 \operatorname{mod} a_n$. This is a chain of congruences; hence, an application of Proposition 2.16 shows that $a_{nm+n} \equiv 0 \operatorname{mod} a_n$. (In the future, we shall no longer explicitly say things like this; we shall leave it to the reader to apply Proposition 2.16 to any chain of congruences that we write down.)

Proposition 2.11 **(a)** (applied to a_{nm+n} and a_n instead of a and n) shows that we have $a_{nm+n} \equiv 0 \operatorname{mod} a_n$ if and only if $a_n \mid a_{nm+n}$. Hence, we have $a_n \mid a_{nm+n}$ (since $a_{nm+n} \equiv 0 \operatorname{mod} a_n$). In view of $nm + n = n(m+1)$, this rewrites as $a_n \mid a_{n(m+1)}$. In other words, (58) holds for $w = m+1$. This completes the induction step. Thus, (58) is proven by induction.

Therefore, Theorem 2.24 **(b)** is proven.

(c) Let u and v be two nonnegative integers satisfying $u \mid v$. We must prove that $a_u \mid a_v$. If $v = 0$, then this is obvious (because if $v = 0$, then $a_v = a_0 = 0 = 0 a_u$ and therefore $a_u \mid a_v$). Hence, for the rest of this proof, we can WLOG assume that we don't have $v = 0$. Assume this.

Thus, we don't have $v = 0$. Hence, $v \neq 0$, so that $v > 0$ (since v is nonnegative).

But u divides v (since $u \mid v$). In other words, there exists an integer w such that $v = uw$. Consider this w. If we had $w < 0$, then we would have $uw \leq 0$ (since u is nonnegative), which would contradict $uw = v > 0$. Hence, we cannot have $w < 0$. Thus, we must have $w \geq 0$. Therefore, $w \in \mathbb{N}$. Hence, Theorem 2.24 **(b)** (applied to $n = u$) yields $a_u \mid a_{uw}$. In view of $v = uw$, this rewrites as $a_u \mid a_v$. This proves Theorem 2.24 **(c)**. ☐

Applying Theorem 2.24 **(c)** to $q = 2$ and $r = 1$, we obtain the observation about divisibility made in Example 2.23.

2.3.2. The Fibonacci sequence and a generalization

Another example of a recursively defined sequence is the famous Fibonacci sequence:

Example 2.25. The *Fibonacci sequence* is the sequence $(f_0, f_1, f_2, \ldots)$ of integers which is defined recursively by

$$f_0 = 0, \qquad f_1 = 1, \qquad \text{and}$$
$$f_n = f_{n-1} + f_{n-2} \qquad \text{for all } n \geq 2.$$

Let us compute its first few entries:

$$\begin{aligned}
f_0 &= 0; \\
f_1 &= 1; \\
f_2 &= \underbrace{f_1}_{=1} + \underbrace{f_0}_{=0} = 1 + 0 = 1; \\
f_3 &= \underbrace{f_2}_{=1} + \underbrace{f_1}_{=1} = 1 + 1 = 2; \\
f_4 &= \underbrace{f_3}_{=2} + \underbrace{f_2}_{=1} = 2 + 1 = 3; \\
f_5 &= \underbrace{f_4}_{=3} + \underbrace{f_3}_{=2} = 3 + 2 = 5; \\
f_6 &= \underbrace{f_5}_{=5} + \underbrace{f_4}_{=3} = 5 + 3 = 8.
\end{aligned}$$

Again, we observe (as in Example 2.23) that $f_2 \mid f_6$ and $f_3 \mid f_6$, which suggests that we might have $f_u \mid f_v$ whenever u and v are two nonnegative integers satisfying $u \mid v$.

Some further experimentation may suggest that the equality $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$ holds for all $n \in \mathbb{N}$ and $m \in \mathbb{N}$.

Both of these conjectures will be shown in the following theorem, in greater generality.

Theorem 2.26. Fix some $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$. Let $(x_0, x_1, x_2, \ldots)$ be a sequence of integers defined recursively by

$$x_0 = 0, \qquad x_1 = 1, \qquad \text{and}$$
$$x_n = a x_{n-1} + b x_{n-2} \qquad \text{for each } n \geq 2.$$

(Note that if $a = 1$ and $b = 1$, then this sequence $(x_0, x_1, x_2, \ldots)$ is precisely the Fibonacci sequence $(f_0, f_1, f_2, \ldots)$ from Example 2.25. If $a = 0$ and $b = 1$, then our sequence $(x_0, x_1, x_2, \ldots)$ is the sequence $(0, 1, 0, b, 0, b^2, 0, b^3, \ldots)$ that alternates between 0's and powers of b. The reader can easily work out further examples.)

(a) We have $x_{n+m+1} = bx_n x_m + x_{n+1} x_{m+1}$ for all $n \in \mathbb{N}$ and $m \in \mathbb{N}$.

(b) For any $n \in \mathbb{N}$ and $w \in \mathbb{N}$, we have $x_n \mid x_{nw}$.

(c) If u and v are two nonnegative integers satisfying $u \mid v$, then $x_u \mid x_v$.

Before we prove this theorem, let us discuss how **not** to prove it:

Remark 2.27. The proof of Theorem 2.26 **(a)** below illustrates an important aspect of induction proofs: Namely, when devising an induction proof, we often have not only a choice of what variable to induct on (e.g., we could try proving Theorem 2.26 **(a)** by induction on n or by induction on m), but also a choice of whether to leave the other variables fixed. For example, let us try to prove Theorem 2.26 **(a)** by induction on n while leaving the variable m fixed. That is, we fix some $m \in \mathbb{N}$, and we define $\mathcal{A}(n)$ (for each $n \in \mathbb{N}$) to be the following statement:

$$(x_{n+m+1} = bx_n x_m + x_{n+1} x_{m+1}).$$

Then, it is easy to check that $\mathcal{A}(0)$ holds, so the induction base is complete. For the induction step, we fix some $k \in \mathbb{N}$. (This k serves the role of the "m" in Theorem 2.1, but we cannot call it m here since m already stands for a fixed number.) We assume that $\mathcal{A}(k)$ holds, and we intend to prove $\mathcal{A}(k+1)$.

Our induction hypothesis says that $\mathcal{A}(k)$ holds; in other words, we have $x_{k+m+1} = bx_k x_m + x_{k+1} x_{m+1}$. We want to prove $\mathcal{A}(k+1)$; in other words, we want to prove that $x_{(k+1)+m+1} = bx_{k+1} x_m + x_{(k+1)+1} x_{m+1}$.

A short moment of deliberation shows that we cannot do this (at least not with our current knowledge). There is no direct way of deriving $\mathcal{A}(k+1)$ from $\mathcal{A}(k)$. **However**, if we knew that the statement $\mathcal{A}(k)$ holds "for $m+1$ instead of m" (that is, if we knew that $x_{k+(m+1)+1} = bx_k x_{m+1} + x_{k+1} x_{(m+1)+1}$), then we could derive $\mathcal{A}(k+1)$. But we cannot just "apply $\mathcal{A}(k)$ to $m+1$ instead of m"; after all, m is a fixed number, so we cannot have it take different values in $\mathcal{A}(k)$ and in $\mathcal{A}(k+1)$.

So we are at an impasse. We got into this impasse by fixing m. So let us try **not** fixing $m \in \mathbb{N}$ right away, but instead defining $\mathcal{A}(n)$ (for each $n \in \mathbb{N}$) to be the following statement:

$$(x_{n+m+1} = bx_n x_m + x_{n+1} x_{m+1} \text{ for all } m \in \mathbb{N}).$$

Thus, $\mathcal{A}(n)$ is not a statement about a specific integer m any more, but rather a statement about all nonnegative integers m. This allows us to apply $\mathcal{A}(k)$ to $m+1$ instead of m in the induction step. (We can still fix $m \in \mathbb{N}$ **during the induction step**; this doesn't prevent us from applying $\mathcal{A}(k)$ to $m+1$ instead of m, since $\mathcal{A}(k)$ has been formulated before m was fixed.) This way, we arrive at the following proof:

Proof of Theorem 2.26. **(a)** We claim that for each $n \in \mathbb{N}$, we have

$$\left(x_{n+m+1} = bx_n x_m + x_{n+1} x_{m+1} \text{ for all } m \in \mathbb{N}\right). \tag{59}$$

Indeed, let us prove (59) by induction on n:

Induction base: We have $x_{0+m+1} = bx_0 x_m + x_{0+1} x_{m+1}$ for all $m \in \mathbb{N}$ [45]. In other words, (59) holds for $n = 0$. This completes the induction base.

Induction step: Let $k \in \mathbb{N}$. Assume that (59) holds for $n = k$. We must prove that (59) holds for $n = k + 1$.

We have assumed that (59) holds for $n = k$. In other words, we have

$$\left(x_{k+m+1} = bx_k x_m + x_{k+1} x_{m+1} \text{ for all } m \in \mathbb{N}\right). \tag{60}$$

Now, let $m \in \mathbb{N}$. We have $m + 2 \geq 2$; thus, the recursive definition of the sequence $(x_0, x_1, x_2, \ldots)$ yields

$$x_{m+2} = a \underbrace{x_{(m+2)-1}}_{=x_{m+1}} + b \underbrace{x_{(m+2)-2}}_{=x_m} = ax_{m+1} + bx_m. \tag{61}$$

The same argument (with m replaced by k) yields

$$x_{k+2} = ax_{k+1} + bx_k. \tag{62}$$

But we can apply (60) to $m + 1$ instead of m. Thus, we obtain

$$\begin{aligned}
x_{k+(m+1)+1} &= bx_k x_{m+1} + x_{k+1} \underbrace{x_{(m+1)+1}}_{\substack{=x_{m+2}=ax_{m+1}+bx_m \\ \text{(by (61))}}} \\
&= bx_k x_{m+1} + \underbrace{x_{k+1}\left(ax_{m+1} + bx_m\right)}_{=ax_{k+1}x_{m+1}+bx_{k+1}x_m} = \underbrace{bx_k x_{m+1} + ax_{k+1} x_{m+1}}_{=(ax_{k+1}+bx_k)x_{m+1}} + bx_{k+1} x_m \\
&= \underbrace{\left(ax_{k+1} + bx_k\right)}_{\substack{=x_{k+2} \\ \text{(by (62))}}} x_{m+1} + bx_{k+1} x_m = x_{k+2} x_{m+1} + bx_{k+1} x_m \\
&= bx_{k+1} x_m + \underbrace{x_{k+2}}_{=x_{(k+1)+1}} x_{m+1} = bx_{k+1} x_m + x_{(k+1)+1} x_{m+1}.
\end{aligned}$$

In view of $k + (m + 1) + 1 = (k + 1) + m + 1$, this rewrites as

$$x_{(k+1)+m+1} = bx_{k+1} x_m + x_{(k+1)+1} x_{m+1}.$$

Now, forget that we fixed m. We thus have shown that $x_{(k+1)+m+1} = bx_{k+1} x_m + x_{(k+1)+1} x_{m+1}$ for all $m \in \mathbb{N}$. In other words, (59) holds for $n = k + 1$. This completes the induction step. Thus, (59) is proven.

[45] *Proof.* Let $m \in \mathbb{N}$. Then, $x_{0+m+1} = x_{m+1}$. Comparing this with $b \underbrace{x_0}_{=0} x_m + \underbrace{x_{0+1}}_{=x_1=1} x_{m+1} = b0x_m + 1x_{m+1} = x_{m+1}$, we obtain $x_{0+m+1} = bx_0 x_m + x_{0+1} x_{m+1}$, qed.

Hence, Theorem 2.26 **(a)** holds.

(b) Fix $n \in \mathbb{N}$. We claim that

$$x_n \mid x_{nw} \qquad \text{for each } w \in \mathbb{N}. \tag{63}$$

Indeed, let us prove (63) by induction on w:

Induction base: We have $x_{n\cdot 0} = x_0 = 0 = 0x_n$ and thus $x_n \mid x_{n\cdot 0}$. In other words, (63) holds for $w = 0$. This completes the induction base.

Induction step: Let $k \in \mathbb{N}$. Assume that (63) holds for $w = k$. We must now prove that (63) holds for $w = k+1$. In other words, we must prove that $x_n \mid x_{n(k+1)}$.

If $n = 0$, then this is true[46]. Hence, for the rest of this proof, we can WLOG assume that we don't have $n = 0$. Assume this.

We have assumed that (63) holds for $w = k$. In other words, we have $x_n \mid x_{nk}$. In other words, $x_{nk} \equiv 0 \mod x_n$. [47] Likewise, from $x_n \mid x_n$, we obtain $x_n \equiv 0 \mod x_n$.

We have $n \in \mathbb{N}$ but $n \neq 0$ (since we don't have $n = 0$). Hence, n is a positive integer. Thus, $n-1 \in \mathbb{N}$. Therefore, Theorem 2.26 **(a)** (applied to nk and $n-1$ instead of n and m) yields

$$x_{nk+(n-1)+1} = bx_{nk}x_{n-1} + x_{nk+1}x_{(n-1)+1}.$$

In view of $nk + (n-1) + 1 = n(k+1)$, this rewrites as

$$x_{n(k+1)} = b \underbrace{x_{nk}}_{\equiv 0 \mod x_n} x_{n-1} + x_{nk+1} \underbrace{x_{(n-1)+1}}_{= x_n \equiv 0 \mod x_n} \equiv b0x_{n-1} + x_{nk+1}0 = 0 \mod x_n.$$

[48] Thus, we have shown that $x_{n(k+1)} \equiv 0 \mod x_n$. In other words, $x_n \mid x_{n(k+1)}$ (again, this follows from Proposition 2.11 **(a)**). In other words, (63) holds for $w = k+1$. This completes the induction step. Hence, (63) is proven by induction.

---

[46] *Proof.* Let us assume that $n = 0$. Then, $x_{n(k+1)} = x_{0(k+1)} = x_0 = 0 = 0x_n$, and thus $x_n \mid x_{n(k+1)}$, qed.

[47] Here, again, we have used Proposition 2.11 **(a)** (applied to x_{nk} and x_n instead of a and n). This argument is simple enough that we will leave it unsaid in the future.

[48] We have used substitutivity for congruences in this computation. Here is, again, a way to rewrite it without this use:

We have $x_{n(k+1)} = bx_{nk}x_{n-1} + x_{nk+1}x_{(n-1)+1}$. But $b \equiv b \mod x_n$ (by Proposition 2.12 **(a)**) and $x_{n-1} \equiv x_{n-1} \mod x_n$ (for the same reason) and $x_{nk+1} \equiv x_{nk+1} \mod x_n$ (for the same reason). Now, Proposition 2.21 **(c)** (applied to b, b, x_{nk}, 0 and x_n instead of a, b, c, d and n) yields $bx_{nk} \equiv b0 \mod x_n$ (since $b \equiv b \mod x_n$ and $x_{nk} \equiv 0 \mod x_n$). Hence, Proposition 2.21 **(c)** (applied to bx_{nk}, $b0$, x_{n-1}, x_{n-1} and x_n instead of a, b, c, d and n) yields $bx_{nk}x_{n-1} \equiv b0x_{n-1} \mod x_n$ (since $bx_{nk} \equiv b0 \mod x_n$ and $x_{n-1} \equiv x_{n-1} \mod x_n$). Also, $x_{nk+1}x_{(n-1)+1} \equiv x_{nk+1}x_{(n-1)+1} \mod x_n$ (by Proposition 2.12 **(a)**). Hence, Proposition 2.21 **(a)** (applied to $bx_{nk}x_{n-1}$, $b0x_{n-1}$, $x_{nk+1}x_{(n-1)+1}$, $x_{nk+1}x_{(n-1)+1}$ and x_n instead of a, b, c, d and n) yields

$$bx_{nk}x_{n-1} + x_{nk+1}x_{(n-1)+1} \equiv b0x_{n-1} + x_{nk+1}x_{(n-1)+1} \mod x_n$$

(since $bx_{nk}x_{n-1} \equiv b0x_{n-1} \mod x_n$ and $x_{nk+1}x_{(n-1)+1} \equiv x_{nk+1}x_{(n-1)+1} \mod x_n$).

Also, Proposition 2.21 **(c)** (applied to x_{nk+1}, x_{nk+1}, $x_{(n-1)+1}$, 0 and x_n instead of a, b, c, d and n) yields $x_{nk+1}x_{(n-1)+1} \equiv x_{nk+1}0 \mod x_n$ (since $x_{nk+1} \equiv x_{nk+1} \mod x_n$ and $x_{(n-1)+1} = x_n \equiv 0 \mod x_n$). Furthermore, $b0x_{n-1} \equiv b0x_{n-1} \mod x_n$ (by Proposition 2.12 **(a)**). Finally, Proposition

This proves Theorem 2.26 **(b)**.

(c) Theorem 2.26 **(c)** can be derived from Theorem 2.26 **(b)** in the same way as Theorem 2.24 **(c)** was derived from Theorem 2.24 **(b)**. □

Applying Theorem 2.26 **(a)** to $a = 1$ and $b = 1$, we obtain the equality $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$ noticed in Example 2.25. Applying Theorem 2.26 **(c)** to $a = 1$ and $b = 1$, we obtain the observation about divisibility made in Example 2.25.

Note that part **(a)** of Theorem 2.26 still works if a and b are real numbers (instead of being integers). But of course, in this case, $(x_0, x_1, x_2, \ldots)$ will be merely a sequence of real numbers (rather than a sequence of integers), and thus parts **(b)** and **(c)** of Theorem 2.26 will no longer make sense (since divisibility is only defined for integers).

2.4. The sum of the first n positive integers

We now come to one of the most classical examples of a proof by induction: Namely, we shall prove the fact that for each $n \in \mathbb{N}$, the sum of the first n positive integers (that is, the sum $1 + 2 + \cdots + n$) equals $\dfrac{n(n+1)}{2}$. However, there is a catch here, which is easy to overlook if one isn't trying to be completely rigorous: We don't really know yet whether there is such a thing as "the sum of the first n positive integers"! To be more precise, we have introduced the $\sum$ sign in Section 1.4, which would allow us to define the sum of the first n positive integers (as $\sum_{i=1}^{n} i$); but our definition of the $\sum$ sign relied on a fact which we have not proved yet (namely, the fact that the right hand side of (1) does not depend on the choice of t). We shall prove this fact later (Theorem 2.118 **(a)**), but for now we prefer not to use it. Instead, let us replace the notion of "the sum of the first n positive integers" by a recursively defined sequence:

Proposition 2.28. Let $(t_0, t_1, t_2, \ldots)$ be a sequence of integers defined recursively by

$$t_0 = 0, \qquad \text{and}$$
$$t_n = t_{n-1} + n \qquad \text{for each } n \geq 1.$$

2.21 **(a)** (applied to $b0x_{n-1}$, $b0x_{n-1}$, $x_{nk+1}x_{(n-1)+1}$, $x_{nk+1}0$ and x_n instead of a, b, c, d and n) yields

$$b0x_{n-1} + x_{nk+1}x_{(n-1)+1} \equiv b0x_{n-1} + x_{nk+1}0 \operatorname{mod} x_n$$

(since $b0x_{n-1} \equiv b0x_{n-1} \operatorname{mod} x_n$ and $x_{nk+1}x_{(n-1)+1} \equiv x_{nk+1}0 \operatorname{mod} x_n$). Thus,

$$\begin{aligned} x_{n(k+1)} &= bx_{nk}x_{n-1} + x_{nk+1}x_{(n-1)+1} \equiv b0x_{n-1} + x_{nk+1}x_{(n-1)+1} \\ &\equiv b0x_{n-1} + x_{nk+1}0 = 0 \operatorname{mod} x_n. \end{aligned}$$

So we have proven that $x_{n(k+1)} \equiv 0 \operatorname{mod} x_n$.

Then,

$$t_n = \frac{n(n+1)}{2} \qquad \text{for each } n \in \mathbb{N}. \tag{64}$$

The sequence $(t_0, t_1, t_2, \ldots)$ defined in Proposition 2.28 is known as the *sequence of triangular numbers*. Its definition shows that

$$\begin{aligned}
t_0 &= 0; \\
t_1 &= \underbrace{t_0}_{=0} + 1 = 0 + 1 = 1; \\
t_2 &= \underbrace{t_1}_{=1} + 2 = 1 + 2; \\
t_3 &= \underbrace{t_2}_{=1+2} + 3 = (1+2) + 3; \\
t_4 &= \underbrace{t_3}_{=(1+2)+3} + 4 = ((1+2)+3) + 4; \\
t_5 &= \underbrace{t_4}_{=((1+2)+3)+4} + 5 = (((1+2)+3)+4) + 5
\end{aligned}$$

[49] and so on; this explains why it makes sense to think of t_n as the sum of the first n positive integers. (This is legitimate even when $n = 0$, because the sum of the first 0 positive integers is an empty sum, and an empty sum is always defined to be equal to 0.) Once we have convinced ourselves that "the sum of the first n positive integers" is a well-defined concept, it will be easy to see (by induction) that t_n **is** the sum of the first n positive integers whenever $n \in \mathbb{N}$. Therefore, Proposition 2.28 will tell us that the sum of the first n positive integers equals $\frac{n(n+1)}{2}$ whenever $n \in \mathbb{N}$.

For now, let us prove Proposition 2.28:

Proof of Proposition 2.28. We shall prove (64) by induction on n:

Induction base: Comparing $t_0 = 0$ with $\frac{0(0+1)}{2} = 0$, we obtain $t_0 = \frac{0(0+1)}{2}$. In other words, (64) holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (64) holds for $n = m$. We must prove that (64) holds for $n = m + 1$.

We have assumed that (64) holds for $n = m$. In other words, we have $t_m = \frac{m(m+1)}{2}$.

[49] Note that we write "$(((1+2)+3)+4)+5$" and not "$1+2+3+4+5$". The reason for this is that we haven't proven yet that the expression "$1+2+3+4+5$" is well-defined. (This expression **is** well-defined, but this will only be clear once we have proven Theorem 2.118 **(a)** below.)

Recall that $t_n = t_{n-1} + n$ for each $n \geq 1$. Applying this to $n = m+1$, we obtain

$$t_{m+1} = \underbrace{t_{(m+1)-1}}_{=t_m=\frac{m(m+1)}{2}} + (m+1) = \frac{m(m+1)}{2} + (m+1) = \frac{m(m+1)+2(m+1)}{2}$$
$$= \frac{(m+2)(m+1)}{2} = \frac{(m+1)(m+2)}{2} = \frac{(m+1)((m+1)+1)}{2}$$

(since $m+2 = (m+1)+1$). In other words, (64) holds for $n = m+1$. This completes the induction step. Hence, (64) is proven by induction. This proves Proposition 2.28. $\square$

2.5. Induction on a derived quantity: maxima of sets

2.5.1. Defining maxima

We have so far been applying the Induction Principle in fairly obvious ways: With the exception of our proof of Proposition 2.16, we have mostly been doing induction on a variable (n or k or i) that already appeared in the claim that we were proving. But sometimes, it is worth doing induction on a variable that does **not** explicitly appear in this claim (which, formally speaking, means that we introduce a new variable to do induction on). For example, the claim might be saying "Each nonempty finite set S of integers has a largest element", and we prove it by induction on $|S| - 1$. This means that instead of directly proving the claim itself, we rather prove the equivalent claim "For each $n \in \mathbb{N}$, each nonempty finite set S of integers satisfying $|S| - 1 = n$ has a largest element" by induction on n. We shall show this proof in more detail below (see Theorem 2.35). First, we prepare by discussing largest elements of sets in general.

Definition 2.29. Let S be a set of integers (or rational numbers, or real numbers). A *maximum* of S is defined to be an element $s \in S$ that satisfies

$$(s \geq t \text{ for each } t \in S).$$

In other words, a maximum of S is defined to be an element of S which is greater or equal to each element of S.

(The plural of the word "maximum" is "maxima".)

Example 2.30. The set $\{2, 4, 5\}$ has exactly one maximum: namely, 5.

The set $\mathbb{N} = \{0, 1, 2, \ldots\}$ has no maximum: If k was a maximum of $\mathbb{N}$, then we would have $k \geq k+1$, which is absurd.

The set $\{0, -1, -2, \ldots\}$ has a maximum: namely, 0.

The set $\varnothing$ has no maximum, since a maximum would have to be an element of $\varnothing$.

In Theorem 2.35, we shall soon show that every nonempty finite set of integers has a maximum. First, we prove that a maximum is unique if it exists:

Proposition 2.31. Let S be a set of integers (or rational numbers, or real numbers). Then, S has **at most one** maximum.

Proof of Proposition 2.31. Let s_1 and s_2 be two maxima of S. We shall show that $s_1 = s_2$.

Indeed, s_1 is a maximum of S. In other words, s_1 is an element $s \in S$ that satisfies $(s \geq t \text{ for each } t \in S)$ (by the definition of a maximum). In other words, s_1 is an element of S and satisfies

$$(s_1 \geq t \text{ for each } t \in S) . \tag{65}$$

The same argument (applied to s_2 instead of s_1) shows that s_2 is an element of S and satisfies

$$(s_2 \geq t \text{ for each } t \in S) . \tag{66}$$

Now, s_1 is an element of S. Hence, (66) (applied to $t = s_1$) yields $s_2 \geq s_1$. But the same argument (with the roles of s_1 and s_2 interchanged) shows that $s_1 \geq s_2$. Combining this with $s_2 \geq s_1$, we obtain $s_1 = s_2$.

Now, forget that we fixed s_1 and s_2. We thus have shown that if s_1 and s_2 are two maxima of S, then $s_1 = s_2$. In other words, any two maxima of S are equal. In other words, S has **at most one** maximum. This proves Proposition 2.31. □

Definition 2.32. Let S be a set of integers (or rational numbers, or real numbers). Proposition 2.31 shows that S has **at most one** maximum. Thus, if S has a maximum, then this maximum is the unique maximum of S; we shall thus call it *the maximum* of S or *the largest element* of S. We shall denote this maximum by $\max S$.

Thus, if S is a set of integers (or rational numbers, or real numbers) that has a maximum, then this maximum $\max S$ satisfies

$$\max S \in S \tag{67}$$

and

$$(\max S \geq t \text{ for each } t \in S) \tag{68}$$

(because of the definition of a maximum).

Let us next show two simple facts:

Lemma 2.33. Let x be an integer (or rational number, or real number). Then, the set $\{x\}$ has a maximum, namely x.

Proof of Lemma 2.33. Clearly, $x \geq x$. Thus, $x \geq t$ for each $t \in \{x\}$ (because the only $t \in \{x\}$ is x). In other words, x is an element $s \in \{x\}$ that satisfies $(s \geq t \text{ for each } t \in \{x\})$ (since $x \in \{x\}$).

But recall that a maximum of $\{x\}$ means an element $s \in \{x\}$ that satisfies $(s \geq t \text{ for each } t \in \{x\})$ (by the definition of a maximum). Hence, x is a maximum of $\{x\}$ (since x is such an element). Thus, the set $\{x\}$ has a maximum, namely x. This proves Lemma 2.33. □

Proposition 2.34. Let P and Q be two sets of integers (or rational numbers, or real numbers). Assume that P has a maximum, and assume that Q has a maximum. Then, the set $P \cup Q$ has a maximum.

Proof of Proposition 2.34. We know that P has a maximum; it is denoted by $\max P$. We also know that Q has a maximum; it is denoted by $\max Q$. The sets P and Q play symmetric roles in Proposition 2.34 (since $P \cup Q = Q \cup P$). Thus, we can WLOG assume that $\max P \geq \max Q$ (since otherwise, we can simply swap P with Q, without altering the meaning of Proposition 2.34). Assume this.

Now, (67) (applied to $S = P$) shows that $\max P \in P \subseteq P \cup Q$. Furthermore, we claim that

$$(\max P \geq t \text{ for each } t \in P \cup Q). \tag{69}$$

[*Proof of (69):* Let $t \in P \cup Q$. We must show that $\max P \geq t$.

We have $t \in P \cup Q$. In other words, $t \in P$ or $t \in Q$. Hence, we are in one of the following two cases:

Case 1: We have $t \in P$.

Case 2: We have $t \in Q$.

(These two cases might have overlap, but there is nothing wrong about this.)

Let us first consider Case 1. In this case, we have $t \in P$. Hence, (68) (applied to $S = P$) yields $\max P \geq t$. Hence, $\max P \geq t$ is proven in Case 1.

Let us next consider Case 2. In this case, we have $t \in Q$. Hence, (68) (applied to $S = Q$) yields $\max Q \geq t$. Hence, $\max P \geq \max Q \geq t$. Thus, $\max P \geq t$ is proven in Case 2.

We have now proven $\max P \geq t$ in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that $\max P \geq t$ always holds. This proves (69).]

Now, $\max P$ is an element $s \in P \cup Q$ that satisfies $(s \geq t \text{ for each } t \in P \cup Q)$ (since $\max P \in P \cup Q$ and $(\max P \geq t \text{ for each } t \in P \cup Q)$).

But recall that a maximum of $P \cup Q$ means an element $s \in P \cup Q$ that satisfies $(s \geq t \text{ for each } t \in P \cup Q)$ (by the definition of a maximum). Hence, $\max P$ is a maximum of $P \cup Q$ (since $\max P$ is such an element). Thus, the set $P \cup Q$ has a maximum. This proves Proposition 2.34. □

2.5.2. Nonempty finite sets of integers have maxima

Theorem 2.35. Let S be a nonempty finite set of integers. Then, S has a maximum.

First proof of Theorem 2.35. First of all, let us forget that we fixed S. So we want to prove that if S is a nonempty finite set of integers, then S has a maximum.

For each $n \in \mathbb{N}$, we let $\mathcal{A}(n)$ be the statement

$$\left(\begin{array}{c} \text{if } S \text{ is a nonempty finite set of integers satisfying } |S| - 1 = n, \\ \text{then } S \text{ has a maximum} \end{array} \right).$$

We claim that $\mathcal{A}(n)$ holds for all $n \in \mathbb{N}$.

Indeed, let us prove this by induction on n:

Induction base: If S is a nonempty finite set of integers satisfying $|S| - 1 = 0$, then S has a maximum[50]. But this is exactly the statement $\mathcal{A}(0)$. Hence, $\mathcal{A}(0)$ holds. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that $\mathcal{A}(m)$ holds. We shall now show that $\mathcal{A}(m+1)$ holds.

We have assumed that $\mathcal{A}(m)$ holds. In other words,

$$\left(\begin{array}{c} \text{if } S \text{ is a nonempty finite set of integers satisfying } |S| - 1 = m, \\ \text{then } S \text{ has a maximum} \end{array} \right) \tag{70}$$

(because this is what the statement $\mathcal{A}(m)$ says).

Now, let S be a nonempty finite set of integers satisfying $|S| - 1 = m + 1$. There exists some $t \in S$ (since S is nonempty). Consider this t. We have $(S \setminus \{t\}) \cup \{t\} = S \cup \{t\} = S$ (since $t \in S$).

From $t \in S$, we obtain $|S \setminus \{t\}| = |S| - 1 = m + 1 > m \geq 0$ (since $m \in \mathbb{N}$). Hence, the set $S \setminus \{t\}$ is nonempty. Furthermore, this set $S \setminus \{t\}$ is finite (since S is finite) and satisfies $|S \setminus \{t\}| - 1 = m$ (since $|S \setminus \{t\}| = m + 1$). Hence, (70) (applied to $S \setminus \{t\}$ instead of S) shows that $S \setminus \{t\}$ has a maximum. Also, Lemma 2.33 (applied to $x = t$) shows that the set $\{t\}$ has a maximum, namely t. Hence, Proposition 2.34 (applied to $P = S \setminus \{t\}$ and $Q = \{t\}$) shows that the set $(S \setminus \{t\}) \cup \{t\}$ has a maximum. Since $(S \setminus \{t\}) \cup \{t\} = S$, this rewrites as follows: The set S has a maximum.

Now, forget that we fixed S. We thus have shown that if S is a nonempty finite set of integers satisfying $|S| - 1 = m + 1$, then S has a maximum. But this is precisely the statement $\mathcal{A}(m+1)$. Hence, we have shown that $\mathcal{A}(m+1)$ holds. This completes the induction step.

Thus, we have proven (by induction) that $\mathcal{A}(n)$ holds for all $n \in \mathbb{N}$. In other words, for all $n \in \mathbb{N}$, the following holds:

$$\left(\begin{array}{c} \text{if } S \text{ is a nonempty finite set of integers satisfying } |S| - 1 = n, \\ \text{then } S \text{ has a maximum} \end{array} \right) \tag{71}$$

[50] *Proof.* Let S be a nonempty finite set of integers satisfying $|S| - 1 = 0$. We must show that S has a maximum.

Indeed, $|S| = 1$ (since $|S| - 1 = 0$). In other words, S is a 1-element set. In other words, $S = \{x\}$ for some integer x. Consider this x. Lemma 2.33 shows that the set $\{x\}$ has a maximum. In other words, the set S has a maximum (since $S = \{x\}$). This completes our proof.

(because this is what $\mathcal{A}(n)$ says).

Now, let S be a nonempty finite set of integers. We shall prove that S has a maximum.

Indeed, $|S| \in \mathbb{N}$ (since S is finite) and $|S| > 0$ (since S is nonempty); hence, $|S| \geq 1$. Thus, $|S| - 1 \geq 0$, so that $|S| - 1 \in \mathbb{N}$. Hence, we can define $n \in \mathbb{N}$ by $n = |S| - 1$. Consider this n. Thus, $|S| - 1 = n$. Hence, (71) shows that S has a maximum. This proves Theorem 2.35. □

2.5.3. Conventions for writing induction proofs on derived quantities

Let us take a closer look at the proof we just gave. The definition of the statement $\mathcal{A}(n)$ was not exactly unmotivated: This statement simply says that Theorem 2.35 holds under the condition that $|S| - 1 = n$. Thus, by introducing $\mathcal{A}(n)$, we have "sliced" Theorem 2.35 into a sequence of statements $\mathcal{A}(0), \mathcal{A}(1), \mathcal{A}(2), \ldots$, which then allowed us to prove these statements by induction on n even though no "n" appeared in Theorem 2.35 itself. This kind of strategy applies to various other problems. Again, we don't need to explicitly define the statement $\mathcal{A}(n)$ if it is simply saying that the claim we are trying to prove (in our case, Theorem 2.35) holds under the condition that $|S| - 1 = n$; we can just say that we are doing "induction on $|S| - 1$". More generally:

Convention 2.36. Let $\mathcal{B}$ be a logical statement that involves some variables $v_1, v_2, v_3, \ldots$. (For example, $\mathcal{B}$ can be the statement of Theorem 2.35; then, there is only one variable, namely S.)

Let q be some expression (involving the variables $v_1, v_2, v_3, \ldots$ or some of them) that has the property that whenever the variables $v_1, v_2, v_3, \ldots$ satisfy the assumptions of $\mathcal{B}$, the expression q evaluates to some nonnegative integer. (For example, if $\mathcal{B}$ is the statement of Theorem 2.35, then q can be the expression $|S| - 1$, because it is easily seen that if S is a nonempty finite set of integers, then $|S| - 1$ is a nonnegative integer.)

Assume that you want to prove the statement $\mathcal{B}$. Then, you can proceed as follows: For each $n \in \mathbb{N}$, define $\mathcal{A}(n)$ to be the statement saying that[51]

$$(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = n).$$

Then, prove $\mathcal{A}(n)$ by induction on n. Thus:

- The *induction base* consists in proving that the statement $\mathcal{B}$ holds under the condition that $q = 0$.
- The *induction step* consists in fixing $m \in \mathbb{N}$, and showing that if the statement $\mathcal{B}$ holds under the condition that $q = m$, then the statement $\mathcal{B}$ holds under the condition that $q = m + 1$.

Once this induction proof is finished, it immediately follows that the statement $\mathcal{B}$ always holds.

This strategy of proof is called "induction on q" (or "induction over q"). Once you have specified what q is, you don't need to explicitly define $\mathcal{A}(n)$, nor do you ever need to mention n.

Using this convention, we can rewrite our above proof of Theorem 2.35 as follows:

First proof of Theorem 2.35 (second version). It is easy to see that $|S| - 1 \in \mathbb{N}$ [52]. Hence, we can apply induction on $|S| - 1$ to prove Theorem 2.35:

Induction base: Theorem 2.35 holds under the condition that $|S| - 1 = 0$ [53]. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.35 holds under the condition that $|S| - 1 = m$. We shall now show that Theorem 2.35 holds under the condition that $|S| - 1 = m + 1$.

We have assumed that Theorem 2.35 holds under the condition that $|S| - 1 = m$. In other words,

$$\left(\begin{array}{c} \text{if } S \text{ is a nonempty finite set of integers satisfying } |S| - 1 = m, \\ \text{then } S \text{ has a maximum} \end{array} \right). \tag{72}$$

Now, let S be a nonempty finite set of integers satisfying $|S| - 1 = m + 1$. There exists some $t \in S$ (since S is nonempty). Consider this t. We have $(S \setminus \{t\}) \cup \{t\} = S \cup \{t\} = S$ (since $t \in S$).

From $t \in S$, we obtain $|S \setminus \{t\}| = |S| - 1 = m + 1 > m \geq 0$ (since $m \in \mathbb{N}$). Hence, the set $S \setminus \{t\}$ is nonempty. Furthermore, this set $S \setminus \{t\}$ is finite (since S is finite) and satisfies $|S \setminus \{t\}| - 1 = m$ (since $|S \setminus \{t\}| = m + 1$). Hence, (72) (applied to $S \setminus \{t\}$ instead of S) shows that $S \setminus \{t\}$ has a maximum. Also, Lemma 2.33 (applied to $x = t$) shows that the set $\{t\}$ has a maximum, namely t. Hence, Proposition 2.34 (applied to $P = S \setminus \{t\}$ and $Q = \{t\}$) shows that the set $(S \setminus \{t\}) \cup \{t\}$ has a maximum. Since $(S \setminus \{t\}) \cup \{t\} = S$, this rewrites as follows: The set S has a maximum.

Now, forget that we fixed S. We thus have shown that if S is a nonempty finite set of integers satisfying $|S| - 1 = m + 1$, then S has a maximum. In other words, Theorem 2.35 holds under the condition that $|S| - 1 = m + 1$. This completes the induction step. Thus, the induction proof of Theorem 2.35 is complete. □

[51] We assume that no variable named "n" appears in the statement $\mathcal{B}$; otherwise, we need a different letter for our new variable in order to avoid confusion.

[52] *Proof.* We have $|S| \in \mathbb{N}$ (since S is finite) and $|S| > 0$ (since S is nonempty); hence, $|S| \geq 1$. Thus, $|S| - 1 \in \mathbb{N}$, qed.

[53] *Proof.* Let S be as in Theorem 2.35, and assume that $|S| - 1 = 0$. We must show that the claim of Theorem 2.35 holds.

Indeed, $|S| = 1$ (since $|S| - 1 = 0$). In other words, S is a 1-element set. In other words, $S = \{x\}$ for some integer x. Consider this x. Lemma 2.33 shows that the set $\{x\}$ has a maximum. In other words, the set S has a maximum (since $S = \{x\}$). In other words, the claim of Theorem 2.35 holds. This completes our proof.

We could have shortened this proof even further if we didn't explicitly state (72), but rather (instead of applying (72)) said that "we can apply Theorem 2.35 to $S \setminus \{t\}$ instead of S".

Let us stress again that, in order to prove Theorem 2.35 by induction on $|S| - 1$, we had to check that $|S| - 1 \in \mathbb{N}$ whenever S satisfies the assumptions of Theorem 2.35.[54] This check was necessary. For example, if we had instead tried to proceed by induction on $|S| - 2$, then we would only have proven Theorem 2.35 under the condition that $|S| - 2 \in \mathbb{N}$; but this condition isn't always satisfied (indeed, it misses the case when S is a 1-element set).

2.5.4. Vacuous truth and induction bases

Can we also prove Theorem 2.35 by induction on $|S|$ (instead of $|S| - 1$)? This seems a bit strange, since $|S|$ can never be 0 in Theorem 2.35 (because S is required to be nonempty), so that the induction base would be talking about a situation that never occurs. However, there is nothing wrong about it, and we already do talk about such situations oftentimes (for example, every time we make a proof by contradiction). The following concept from basic logic explains this:

Convention 2.37. (a) A logical statement of the form "if $\mathcal{A}$, then $\mathcal{B}$" (where $\mathcal{A}$ and $\mathcal{B}$ are two statements) is said to be *vacuously true* if $\mathcal{A}$ does not hold. For example, the statement "if $0 = 1$, then every set is empty" is vacuously true, because $0 = 1$ is false. The statement "if $0 = 1$, then $1 = 1$" is also vacuously true, although its truth can also be seen as a consequence of the fact that $1 = 1$ is true.

By the laws of logic, a vacuously true statement is always true! This may sound counterintuitive, but actually makes perfect sense: A statement "if $\mathcal{A}$, then $\mathcal{B}$" only says anything about situations where $\mathcal{A}$ holds. If $\mathcal{A}$ never holds, then it therefore says nothing. And when you are saying nothing, you are certainly not lying.

The principle that a vacuously true statement always holds is known as "*ex falso quodlibet*" (literal translation: "from the false, anything") or "*principle of explosion*". It can be restated as follows: From a false statement, any statement follows.

(b) Now, let X be a set, and let $\mathcal{A}(x)$ and $\mathcal{B}(x)$ be two statements defined for each $x \in X$. A statement of the form "for each $x \in X$ satisfying $\mathcal{A}(x)$, we have $\mathcal{B}(x)$" will automatically hold if there exists no $x \in X$ satisfying $\mathcal{A}(x)$. (Indeed, this statement can be rewritten as "for each $x \in X$, we have (if $\mathcal{A}(x)$, then $\mathcal{B}(x)$)"; but this holds because the statement "if $\mathcal{A}(x)$, then $\mathcal{B}(x)$" is vacuously true for each $x \in X$.) Such a statement will also be called *vacuously true*.

[54] In our first version of the above proof, we checked this at the end; in the second version, we checked it at the beginning of the proof.

For example, the statement "if $n \in \mathbb{N}$ is both odd and even, then $n = n + 1$" is vacuously true, since no $n \in \mathbb{N}$ can be both odd and even at the same time.

(c) Now, let X be the empty set (that is, $X = \varnothing$), and let $\mathcal{B}(x)$ be a statement defined for each $x \in X$. Then, a statement of the form "for each $x \in X$, we have $\mathcal{B}(x)$" will automatically hold. (Indeed, this statement can be rewritten as "for each $x \in X$, we have (if $x \in X$, then $\mathcal{B}(x)$)"; but this holds because the statement "if $x \in X$, then $\mathcal{B}(x)$" is vacuously true for each $x \in X$, since its premise ($x \in X$) is false.) Again, such a statement is said to be *vacuously true*.

For example, the statement "for each $x \in \varnothing$, we have $x \neq x$" is vacuously true (because there exists no $x \in \varnothing$).

Thus, if we try to prove Theorem 2.35 by induction on $|S|$, then the induction base becomes vacuously true. However, the induction step becomes more complicated, since we can no longer argue that $S \setminus \{t\}$ is nonempty, but instead have to account for the case when $S \setminus \{t\}$ is empty as well. So we gain and we lose at the same time. Here is how this proof looks like:

Second proof of Theorem 2.35. Clearly, $|S| \in \mathbb{N}$ (since S is a finite set). Hence, we can apply induction on $|S|$ to prove Theorem 2.35:

Induction base: Theorem 2.35 holds under the condition that $|S| = 0$ [55]. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.35 holds under the condition that $|S| = m$. We shall now show that Theorem 2.35 holds under the condition that $|S| = m + 1$.

We have assumed that Theorem 2.35 holds under the condition that $|S| = m$. In other words,

$$\left(\begin{array}{c} \text{if } S \text{ is a nonempty finite set of integers satisfying } |S| = m, \\ \text{then } S \text{ has a maximum} \end{array} \right). \tag{73}$$

Now, let S be a nonempty finite set of integers satisfying $|S| = m + 1$. We want to prove that S has a maximum.

There exists some $t \in S$ (since S is nonempty). Consider this t. We have $(S \setminus \{t\}) \cup \{t\} = S \cup \{t\} = S$ (since $t \in S$). Lemma 2.33 (applied to $x = t$) shows that the set $\{t\}$ has a maximum, namely t.

We are in one of the following two cases:

Case 1: We have $S \setminus \{t\} = \varnothing$.

Case 2: We have $S \setminus \{t\} \neq \varnothing$.

Let us first consider Case 1. In this case, we have $S \setminus \{t\} = \varnothing$. Hence, $S \subseteq \{t\}$. Thus, either $S = \varnothing$ or $S = \{t\}$ (since the only subsets of $\{t\}$ are $\varnothing$ and $\{t\}$). Since

[55] *Proof.* Let S be as in Theorem 2.35, and assume that $|S| = 0$. We must show that the claim of Theorem 2.35 holds.

Indeed, $|S| = 0$, so that S is the empty set. This contradicts the assumption that S be nonempty. From this contradiction, we conclude that everything holds (by the "ex falso quodlibet" principle). Thus, in particular, the claim of Theorem 2.35 holds. This completes our proof.

$S = \varnothing$ is impossible (because S is nonempty), we thus have $S = \{t\}$. But the set $\{t\}$ has a maximum. In view of $S = \{t\}$, this rewrites as follows: The set S has a maximum. Thus, our goal (to prove that S has a maximum) is achieved in Case 1.

Let us now consider Case 2. In this case, we have $S \setminus \{t\} \neq \varnothing$. Hence, the set $S \setminus \{t\}$ is nonempty. From $t \in S$, we obtain $|S \setminus \{t\}| = |S| - 1 = m$ (since $|S| = m+1$). Furthermore, the set $S \setminus \{t\}$ is finite (since S is finite). Hence, (73) (applied to $S \setminus \{t\}$ instead of S) shows that $S \setminus \{t\}$ has a maximum. Also, recall that the set $\{t\}$ has a maximum. Hence, Proposition 2.34 (applied to $P = S \setminus \{t\}$ and $Q = \{t\}$) shows that the set $(S \setminus \{t\}) \cup \{t\}$ has a maximum. Since $(S \setminus \{t\}) \cup \{t\} = S$, this rewrites as follows: The set S has a maximum. Hence, our goal (to prove that S has a maximum) is achieved in Case 2.

We have now proven that S has a maximum in each of the two Cases 1 and 2. Therefore, S always has a maximum (since Cases 1 and 2 cover all possibilities).

Now, forget that we fixed S. We thus have shown that if S is a nonempty finite set of integers satisfying $|S| = m+1$, then S has a maximum. In other words, Theorem 2.35 holds under the condition that $|S| = m+1$. This completes the induction step. Thus, the induction proof of Theorem 2.35 is complete. □

2.5.5. Further results on maxima and minima

We can replace "integers" by "rational numbers" or "real numbers" in Theorem 2.35; all the proofs given above still apply then. Thus, we obtain the following:

Theorem 2.38. Let S be a nonempty finite set of integers (or rational numbers, or real numbers). Then, S has a maximum.

Hence, if S is a nonempty finite set of integers (or rational numbers, or real numbers), then $\max S$ is well-defined (because Theorem 2.38 shows that S has a maximum, and Proposition 2.31 shows that this maximum is unique).

Moreover, just as we have defined maxima (i.e., largest elements) of sets, we can define minima (i.e., smallest elements) of sets, and prove similar results about them:

Definition 2.39. Let S be a set of integers (or rational numbers, or real numbers). A *minimum* of S is defined to be an element $s \in S$ that satisfies

$$(s \leq t \text{ for each } t \in S).$$

In other words, a minimum of S is defined to be an element of S which is less or equal to each element of S.

(The plural of the word "minimum" is "minima".)

Example 2.40. The set $\{2,4,5\}$ has exactly one minimum: namely, 2.

The set $\mathbb{N} = \{0,1,2,\ldots\}$ has exactly one minimum: namely, 0.

The set $\{0, -1, -2, \ldots\}$ has no minimum: If k was a minimum of this set, then we would have $k \leq k - 1$, which is absurd.

The set $\varnothing$ has no minimum, since a minimum would have to be an element of $\varnothing$.

The analogue of Proposition 2.31 for minima instead of maxima looks exactly as one would expect it:

Proposition 2.41. Let S be a set of integers (or rational numbers, or real numbers). Then, S has **at most one** minimum.

Proof of Proposition 2.41. To obtain a proof of Proposition 2.41, it suffices to replace every "$\geq$" sign by a "$\leq$" sign (and every word "maximum" by "minimum") in the proof of Proposition 2.31 given above. □

Definition 2.42. Let S be a set of integers (or rational numbers, or real numbers). Proposition 2.41 shows that S has **at most one** minimum. Thus, if S has a minimum, then this minimum is the unique minimum of S; we shall thus call it *the minimum* of S or *the smallest element* of S. We shall denote this minimum by $\min S$.

The analogue of Theorem 2.38 is the following:

Theorem 2.43. Let S be a nonempty finite set of integers (or rational numbers, or real numbers). Then, S has a minimum.

Proof of Theorem 2.43. To obtain a proof of Theorem 2.43, it suffices to replace every "$\geq$" sign by a "$\leq$" sign (and every word "maximum" by "minimum") in the proof of Theorem 2.38 given above (and also in the proofs of all the auxiliary results that were used in said proof).[56] □

Alternatively, Theorem 2.43 can be obtained from Theorem 2.38 by applying the latter theorem to the set $\{-s \mid s \in S\}$. In fact, it is easy to see that a number x is the minimum of S if and only if $-x$ is the maximum of the set $\{-s \mid s \in S\}$. We leave the details of this simple argument to the reader.

We also should mention that Theorem 2.43 holds **without** requiring that S be finite, if we instead require that S consist of nonnegative integers:

Theorem 2.44. Let S be a nonempty set of nonnegative integers. Then, S has a minimum.

But S does not necessarily have a maximum in this situation; the nonnegativity requirement has "broken the symmetry" between maxima and minima.

[56] To be technically precise: not every "$\geq$" sign, of course. The "$\geq$" sign in "$m \geq 0$" should stay unchanged.

We note that the word "integers" is crucial in Theorem 2.44. If we replaced "integers" by "rational numbers", then the theorem would no longer hold (for example, the set of all positive rational numbers has no minimum, since positive rational numbers can get arbitrarily close to 0 yet cannot equal 0).

Proof of Theorem 2.44. The set S is nonempty. Thus, there exists some $p \in S$. Consider this p.

We have $p \in S \subseteq \mathbb{N}$ (since S is a set of nonnegative integers). Thus, $p \in \{0, 1, \ldots, p\}$. Combining this with $p \in S$, we obtain $p \in \{0, 1, \ldots, p\} \cap S$. Hence, the set $\{0, 1, \ldots, p\} \cap S$ contains the element p, and thus is nonempty. Moreover, this set $\{0, 1, \ldots, p\} \cap S$ is a subset of the finite set $\{0, 1, \ldots, p\}$, and thus is finite.

Now we know that $\{0, 1, \ldots, p\} \cap S$ is a nonempty finite set of integers. Hence, Theorem 2.43 (applied to $\{0, 1, \ldots, p\} \cap S$ instead of S) shows that the set $\{0, 1, \ldots, p\} \cap S$ has a minimum. Denote this minimum by m.

Hence, m is a minimum of the set $\{0, 1, \ldots, p\} \cap S$. In other words, m is an element $s \in \{0, 1, \ldots, p\} \cap S$ that satisfies

$$(s \leq t \text{ for each } t \in \{0, 1, \ldots, p\} \cap S)$$

(by the definition of a minimum). In other words, m is an element of $\{0, 1, \ldots, p\} \cap S$ and satisfies

$$(m \leq t \text{ for each } t \in \{0, 1, \ldots, p\} \cap S). \tag{74}$$

Hence, $m \in \{0, 1, \ldots, p\} \cap S \subseteq \{0, 1, \ldots, p\}$, so that $m \leq p$.

Furthermore, $m \in \{0, 1, \ldots, p\} \cap S \subseteq S$. Moreover, we have

$$(m \leq t \text{ for each } t \in S). \tag{75}$$

[*Proof of (75):* Let $t \in S$. We must prove that $m \leq t$.

If $t \in \{0, 1, \ldots, p\} \cap S$, then this follows from (74). Hence, for the rest of this proof, we can WLOG assume that we don't have $t \in \{0, 1, \ldots, p\} \cap S$. Assume this. Thus, $t \notin \{0, 1, \ldots, p\} \cap S$. Combining $t \in S$ with $t \notin \{0, 1, \ldots, p\} \cap S$, we obtain

$$t \in S \setminus (\{0, 1, \ldots, p\} \cap S) = S \setminus \{0, 1, \ldots, p\}.$$

Hence, $t \notin \{0, 1, \ldots, p\}$, so that $t > p$ (since $t \in \mathbb{N}$). Therefore, $t \geq p \geq m$ (since $m \leq p$), so that $m \leq t$. This completes the proof of (75).]

Now, we know that m is an element of S (since $m \in S$) and satisfies $(m \leq t \text{ for each } t \in S)$ (by (75)). In other words, m is an $s \in S$ that satisfies $(s \leq t \text{ for each } t \in S)$. In other words, m is a minimum of S (by the definition of a minimum). Thus, S has a minimum (namely, m). This proves Theorem 2.44. □

2.6. Increasing lists of finite sets

We shall next study (again using induction) another basic feature of finite sets.

We recall that "list" is just a synonym for "tuple"; i.e., a list is a k-tuple for some $k \in \mathbb{N}$. Note that tuples and lists are always understood to be finite.

Definition 2.45. Let S be a set of integers. An *increasing list* of S shall mean a list $(s_1, s_2, \ldots, s_k)$ of elements of S such that $S = \{s_1, s_2, \ldots, s_k\}$ and $s_1 < s_2 < \cdots < s_k$.

In other words, if S is a set of integers, then an increasing list of S means a list such that

- the set S consists of all elements of this list, and
- the elements of this list are strictly increasing.

For example, $(2, 4, 6)$ is an increasing list of the set $\{2, 4, 6\}$, but neither $(2, 6)$ nor $(2, 4, 4, 6)$ nor $(4, 2, 6)$ nor $(2, 4, 5, 6)$ is an increasing list of this set. For another example, $(1, 4, 9, 16)$ is an increasing list of the set $\{i^2 \mid i \in \{1, 2, 3, 4\}\} = \{1, 4, 9, 16\}$. For yet another example, the empty list $()$ is an increasing list of the empty set $\varnothing$.

Now, it is intuitively obvious that any finite set S of integers has a unique increasing list – we just need to list all the elements of S in increasing order, with no repetitions. But from the viewpoint of rigorous mathematics, this needs to be proven. Let us state this as a theorem:

Theorem 2.46. Let S be a finite set of integers. Then, S has exactly one increasing list.

Before we prove this theorem, let us show some auxiliary facts:

Proposition 2.47. Let S be a set of integers. Let $(s_1, s_2, \ldots, s_k)$ be an increasing list of S. Then:

(a) The set S is finite.

(b) We have $|S| = k$.

(c) The elements $s_1, s_2, \ldots, s_k$ are distinct.

Proof of Proposition 2.47. We know that $(s_1, s_2, \ldots, s_k)$ is an increasing list of S. In other words, $(s_1, s_2, \ldots, s_k)$ is a list of elements of S such that $S = \{s_1, s_2, \ldots, s_k\}$ and $s_1 < s_2 < \cdots < s_k$ (by the definition of an "increasing list").

From $S = \{s_1, s_2, \ldots, s_k\}$, we conclude that the set S has at most k elements. Thus, the set S is finite. This proves Proposition 2.47 **(a)**.

We have $s_1 < s_2 < \cdots < s_k$. Hence, if u and v are two elements of $\{1, 2, \ldots, k\}$ such that $u < v$, then $s_u < s_v$ (by Corollary 2.20, applied to $a_i = s_i$) and therefore $s_u \neq s_v$. In other words, the elements $s_1, s_2, \ldots, s_k$ are distinct. This proves Proposition 2.47 **(c)**.

The k elements $s_1, s_2, \ldots, s_k$ are distinct; thus, the set $\{s_1, s_2, \ldots, s_k\}$ has size k. In other words, the set S has size k (since $S = \{s_1, s_2, \ldots, s_k\}$). In other words, $|S| = k$. This proves Proposition 2.47 **(b)**. □

Proposition 2.48. The set $\varnothing$ has exactly one increasing list: namely, the empty list $()$.

Proof of Proposition 2.48. The empty list $()$ satisfies $\varnothing = \{\}$. Thus, the empty list $()$ is a list $(s_1, s_2, \ldots, s_k)$ of elements of $\varnothing$ such that $\varnothing = \{s_1, s_2, \ldots, s_k\}$ and $s_1 < s_2 < \cdots < s_k$ (indeed, the chain of inequalities $s_1 < s_2 < \cdots < s_k$ is vacuously true for the empty list $()$, because it contains no inequality signs). In other words, the empty list $()$ is an increasing list of $\varnothing$ (by the definition of an increasing list). It remains to show that it is the only increasing list of $\varnothing$.

Let $(s_1, s_2, \ldots, s_k)$ be any increasing list of $\varnothing$. Then, Proposition 2.47 **(b)** (applied to $S = \varnothing$) yields $|\varnothing| = k$. Hence, $k = |\varnothing| = 0$, so that $(s_1, s_2, \ldots, s_k) = (s_1, s_2, \ldots, s_0) = ()$.

Now, forget that we fixed $(s_1, s_2, \ldots, s_k)$. We thus have shown that if $(s_1, s_2, \ldots, s_k)$ is any increasing list of $\varnothing$, then $(s_1, s_2, \ldots, s_k) = ()$. In other words, any increasing list of $\varnothing$ is $()$. Therefore, the set $\varnothing$ has exactly one increasing list: namely, the empty list $()$ (since we already know that $()$ is an increasing list of $\varnothing$). This proves Proposition 2.48. ☐

Proposition 2.49. Let S be a nonempty finite set of integers. Let $m = \max S$. Let $(s_1, s_2, \ldots, s_k)$ be any increasing list of S. Then:

(a) We have $k \geq 1$ and $s_k = m$.

(b) The list $(s_1, s_2, \ldots, s_{k-1})$ is an increasing list of $S \setminus \{m\}$.

Proof of Proposition 2.49. We know that $(s_1, s_2, \ldots, s_k)$ is an increasing list of S. In other words, $(s_1, s_2, \ldots, s_k)$ is a list of elements of S such that $S = \{s_1, s_2, \ldots, s_k\}$ and $s_1 < s_2 < \cdots < s_k$ (by the definition of an "increasing list").

Proposition 2.47 **(b)** yields $|S| = k$. Hence, $k = |S| > 0$ (since S is nonempty). Thus, $k \geq 1$ (since k is an integer). Therefore, s_k is well-defined. Clearly, $k \in \{1, 2, \ldots, k\}$ (since $k \geq 1$), so that $s_k \in \{s_1, s_2, \ldots, s_k\} = S$.

We have $s_1 < s_2 < \cdots < s_k$ and thus $s_1 \leq s_2 \leq \cdots \leq s_k$. Hence, Proposition 2.18 (applied to $a_i = s_i$) shows that if u and v are two elements of $\{1, 2, \ldots, k\}$ such that $u \leq v$, then

$$s_u \leq s_v. \tag{76}$$

Thus, we have $(s_k \geq t \text{ for each } t \in S)$ [57]. Hence, s_k is an element $s \in S$ that satisfies $(s \geq t \text{ for each } t \in S)$ (since $s_k \in S$). In other words, s_k is a maximum of S (by the definition of a maximum). Since we know that S has at most one maximum (by Proposition 2.31), we thus conclude that s_k is **the** maximum of S. In other words, $s_k = \max S$. Hence, $s_k = \max S = m$. This completes the proof of Proposition 2.49 **(a)**.

(b) From $s_1 < s_2 < \cdots < s_k$, we obtain $s_1 < s_2 < \cdots < s_{k-1}$. Furthermore, the elements $s_1, s_2, \ldots, s_k$ are distinct (according to Proposition 2.47 **(c)**). In other

[57] *Proof.* Let $t \in S$. Thus, $t \in S = \{s_1, s_2, \ldots, s_k\}$. Hence, $t = s_u$ for some $u \in \{1, 2, \ldots, k\}$. Consider this u. Now, u and k are elements of $\{1, 2, \ldots, k\}$ such that $u \leq k$ (since $u \in \{1, 2, \ldots, k\}$). Hence, (76) (applied to $v = k$) yields $s_u \leq s_k$. Hence, $s_k \geq s_u = t$ (since $t = s_u$), qed.

words, for any two distinct elements u and v of $\{1, 2, \ldots, k\}$, we have

$$s_u \neq s_v. \tag{77}$$

Hence, $s_k \notin \{s_1, s_2, \ldots, s_{k-1}\}$ [58]. Now,

$$\underbrace{S}_{\substack{=\{s_1,s_2,\ldots,s_k\} \\ =\{s_1,s_2,\ldots,s_{k-1}\}\cup\{s_k\}}} \setminus \left\{ \underbrace{m}_{=s_k} \right\} = (\{s_1, s_2, \ldots, s_{k-1}\} \cup \{s_k\}) \setminus \{s_k\}$$
$$= \{s_1, s_2, \ldots, s_{k-1}\} \setminus \{s_k\} = \{s_1, s_2, \ldots, s_{k-1}\}$$

(since $s_k \notin \{s_1, s_2, \ldots, s_{k-1}\}$). Hence, the elements $s_1, s_2, \ldots, s_{k-1}$ belong to the set $S \setminus \{m\}$ (since they clearly belong to the set $\{s_1, s_2, \ldots, s_{k-1}\} = S \setminus \{m\}$). In other words, $(s_1, s_2, \ldots, s_{k-1})$ is a list of elements of $S \setminus \{m\}$.

Now, we know that $(s_1, s_2, \ldots, s_{k-1})$ is a list of elements of $S \setminus \{m\}$ such that $S \setminus \{m\} = \{s_1, s_2, \ldots, s_{k-1}\}$ and $s_1 < s_2 < \cdots < s_{k-1}$. In other words, $(s_1, s_2, \ldots, s_{k-1})$ is an increasing list of $S \setminus \{m\}$. This proves Proposition 2.49 **(b)**. □

We are now ready to prove Theorem 2.46:

Proof of Theorem 2.46. We shall prove Theorem 2.46 by induction on $|S|$:

Induction base: Theorem 2.46 holds under the condition that $|S| = 0$ [59]. This completes the induction base.

Induction step: Let $g \in \mathbb{N}$. Assume that Theorem 2.46 holds under the condition that $|S| = g$. We shall now show that Theorem 2.46 holds under the condition that $|S| = g + 1$.

We have assumed that Theorem 2.46 holds under the condition that $|S| = g$. In other words,

$$\left(\begin{array}{c} \text{if } S \text{ is a finite set of integers satisfying } |S| = g, \\ \text{then } S \text{ has exactly one increasing list} \end{array} \right). \tag{78}$$

Now, let S be a finite set of integers satisfying $|S| = g + 1$. We want to prove that S has exactly one increasing list.

The set S is nonempty (since $|S| = g + 1 > g \geq 0$). Thus, S has a maximum (by Theorem 2.35). Hence, $\max S$ is well-defined. Set $m = \max S$. Thus, $m = \max S \in S$

[58] *Proof.* Assume the contrary. Thus, $s_k \in \{s_1, s_2, \ldots, s_{k-1}\}$. In other words, $s_k = s_u$ for some $u \in \{1, 2, \ldots, k-1\}$. Consider this u. We have $u \in \{1, 2, \ldots, k-1\} \subseteq \{1, 2, \ldots, k\}$.

Now, $u \in \{1, 2, \ldots, k-1\}$, so that $u \leq k - 1 < k$ and thus $u \neq k$. Hence, the elements u and k of $\{1, 2, \ldots, k\}$ are distinct. Thus, (77) (applied to $v = k$) yields $s_u \neq s_k = s_u$. This is absurd. This contradiction shows that our assumption was wrong, qed.

[59] *Proof.* Let S be as in Theorem 2.46, and assume that $|S| = 0$. We must show that the claim of Theorem 2.46 holds.

Indeed, $|S| = 0$, so that S is the empty set. Thus, $S = \varnothing$. But Proposition 2.48 shows that the set $\varnothing$ has exactly one increasing list. Since $S = \varnothing$, this rewrites as follows: The set S has exactly one increasing list. Thus, the claim of Theorem 2.46 holds. This completes our proof.

(by (67)). Therefore, $|S \setminus \{m\}| = |S| - 1 = g$ (since $|S| = g + 1$). Hence, (78) (applied to $S \setminus \{m\}$ instead of S) shows that $S \setminus \{m\}$ has exactly one increasing list. Let $(t_1, t_2, \ldots, t_j)$ be this list. We extend this list to a $(j+1)$-tuple $(t_1, t_2, \ldots, t_{j+1})$ by setting $t_{j+1} = m$.

We have defined $(t_1, t_2, \ldots, t_j)$ as an increasing list of the set $S \setminus \{m\}$. In other words, $(t_1, t_2, \ldots, t_j)$ is a list of elements of $S \setminus \{m\}$ such that $S \setminus \{m\} = \{t_1, t_2, \ldots, t_j\}$ and $t_1 < t_2 < \cdots < t_j$ (by the definition of an "increasing list").

We claim that

$$t_1 < t_2 < \cdots < t_{j+1}. \tag{79}$$

[*Proof of (79):* If $j + 1 \leq 1$, then the chain of inequalities (79) is vacuously true (since it contains no inequality signs). Thus, for the rest of this proof of (79), we WLOG assume that we don't have $j + 1 \leq 1$. Hence, $j + 1 > 1$, so that $j > 0$ and thus $j \geq 1$ (since j is an integer). Hence, t_j is well-defined. We have $j \in \{1, 2, \ldots, j\}$ (since $j \geq 1$) and thus $t_j \in \{t_1, t_2, \ldots, t_j\} = S \setminus \{m\} \subseteq S$. Hence, (68) (applied to $t = t_j$) yields $\max S \geq t_j$. Hence, $t_j \leq \max S = m$. Moreover, $t_j \notin \{m\}$ (since $t_j \in S \setminus \{m\}$); in other words, $t_j \neq m$. Combining this with $t_j \leq m$, we obtain $t_j < m = t_{j+1}$. Combining the chain of inequalities $t_1 < t_2 < \cdots < t_j$ with the single inequality $t_j < t_{j+1}$, we obtain the longer chain of inequalities $t_1 < t_2 < \cdots < t_j < t_{j+1}$. In other words, $t_1 < t_2 < \cdots < t_{j+1}$. This proves (79).]

Next, we shall prove that

$$S = \{t_1, t_2, \ldots, t_{j+1}\}. \tag{80}$$

[*Proof of (80):* We have $(S \setminus \{m\}) \cup \{m\} = S \cup \{m\} = S$ (since $m \in S$). Thus,

$$S = \Bigg(\underbrace{S \setminus \{m\}}_{=\{t_1, t_2, \ldots, t_j\}} \Bigg) \cup \Bigg\{ \underbrace{m}_{=t_{j+1}} \Bigg\} = \{t_1, t_2, \ldots, t_j\} \cup \{t_{j+1}\} = \{t_1, t_2, \ldots, t_j, t_{j+1}\}$$
$$= \{t_1, t_2, \ldots, t_{j+1}\}.$$

This proves (80).]

Clearly, $t_1, t_2, \ldots, t_{j+1}$ are elements of the set $\{t_1, t_2, \ldots, t_{j+1}\}$. In other words, $t_1, t_2, \ldots, t_{j+1}$ are elements of the set S (since $S = \{t_1, t_2, \ldots, t_{j+1}\}$).

Hence, $(t_1, t_2, \ldots, t_{j+1})$ is a list of elements of S. Thus, $(t_1, t_2, \ldots, t_{j+1})$ is a list of elements of S such that $S = \{t_1, t_2, \ldots, t_{j+1}\}$ (by (80)) and $t_1 < t_2 < \cdots < t_{j+1}$ (by (79)). In other words, $(t_1, t_2, \ldots, t_{j+1})$ is an increasing list of S (by the definition of an "increasing list"). Hence, the set S has **at least** one increasing list (namely, $(t_1, t_2, \ldots, t_{j+1})$).

We shall next show that $(t_1, t_2, \ldots, t_{j+1})$ is the only increasing list of S. Indeed, let $(s_1, s_2, \ldots, s_k)$ be any increasing list of S. Then, Proposition 2.49 **(a)** shows that $k \geq 1$ and $s_k = m$. Also, Proposition 2.49 **(b)** shows that the list $(s_1, s_2, \ldots, s_{k-1})$ is an increasing list of $S \setminus \{m\}$.

But recall that $S \setminus \{m\}$ has exactly one increasing list. Thus, in particular, $S \setminus \{m\}$ has **at most** one increasing list. In other words, any two increasing lists of $S \setminus \{m\}$ are equal. Hence, the lists $(s_1, s_2, \ldots, s_{k-1})$ and $(t_1, t_2, \ldots, t_j)$ must be equal (since both of these lists are increasing lists of $S \setminus \{m\}$). In other words, $(s_1, s_2, \ldots, s_{k-1}) = (t_1, t_2, \ldots, t_j)$. In other words, $k - 1 = j$ and

$$(s_i = t_i \text{ for each } i \in \{1, 2, \ldots, k-1\}). \tag{81}$$

From $k - 1 = j$, we obtain $k = j + 1$. Hence, $t_k = t_{j+1} = m$. Next, we claim that

$$s_i = t_i \text{ for each } i \in \{1, 2, \ldots, k\}. \tag{82}$$

[*Proof of (82):* Let $i \in \{1, 2, \ldots, k\}$. We must prove that $s_i = t_i$. If $i \in \{1, 2, \ldots, k-1\}$, then this follows from (81). Hence, for the rest of this proof, we WLOG assume that we don't have $i \in \{1, 2, \ldots, k-1\}$. Hence, $i \notin \{1, 2, \ldots, k-1\}$. Combining $i \in \{1, 2, \ldots, k\}$ with $i \notin \{1, 2, \ldots, k-1\}$, we obtain

$$i \in \{1, 2, \ldots, k\} \setminus \{1, 2, \ldots, k-1\} = \{k\}.$$

In other words, $i = k$. Hence, $s_i = s_k = m = t_k$ (since $t_k = m$). In view of $k = i$, this rewrites as $s_i = t_i$. This proves (82).]

From (82), we obtain $(s_1, s_2, \ldots, s_k) = (t_1, t_2, \ldots, t_k) = (t_1, t_2, \ldots, t_{j+1})$ (since $k = j + 1$).

Now, forget that we fixed $(s_1, s_2, \ldots, s_k)$. We thus have proven that if $(s_1, s_2, \ldots, s_k)$ is any increasing list of S, then $(s_1, s_2, \ldots, s_k) = (t_1, t_2, \ldots, t_{j+1})$. In other words, any increasing list of S equals $(t_1, t_2, \ldots, t_{j+1})$. Thus, the set S has **at most** one increasing list. Since we also know that the set S has **at least** one increasing list, we thus conclude that S has exactly one increasing list.

Now, forget that we fixed S. We thus have shown that

$$\left(\begin{array}{c}\text{if } S \text{ is a finite set of integers satisfying } |S| = g+1, \\ \text{then } S \text{ has exactly one increasing list}\end{array}\right).$$

In other words, Theorem 2.46 holds under the condition that $|S| = g + 1$. This completes the induction step. Hence, Theorem 2.46 is proven by induction. □

Definition 2.50. Let S be a finite set of integers. Theorem 2.46 shows that S has exactly one increasing list. This increasing list is called *the increasing list* of S. It is also called *the list of all elements of S in increasing order (with no repetitions)*. (The latter name, of course, is descriptive.)

The increasing list of S has length $|S|$. (Indeed, if we denote this increasing list by $(s_1, s_2, \ldots, s_k)$, then its length is $k = |S|$, because Proposition 2.47 **(b)** shows that $|S| = k$.)

For each $j \in \{1, 2, \ldots, |S|\}$, we define the *$j$-th smallest element of S* to be the j-th entry of the increasing list of S. In other words, if $(s_1, s_2, \ldots, s_k)$ is the increasing list of S, then the j-th smallest element of S is s_j. Some say "j-th lowest element of S" instead of "j-th smallest element of S".

Remark 2.51. **(a)** Clearly, we can replace the word "integer" by "rational number" or by "real number" in Proposition 2.18, Corollary 2.19, Corollary 2.20, Definition 2.45, Theorem 2.46, Proposition 2.47, Proposition 2.48, Proposition 2.49 and Definition 2.50, because we have not used any properties specific to integers.

(b) If we replace all the "$<$" signs in Definition 2.45 by "$>$" signs, then we obtain the notion of a *decreasing list* of S. There are straightforward analogues of Theorem 2.46, Proposition 2.47, Proposition 2.48 and Proposition 2.49 for decreasing lists (where, of course, the analogue of Proposition 2.49 uses $\min S$ instead of $\max S$). Thus, we can state an analogue of Definition 2.50 as well. In this analogue, the word "increasing" is replaced by "decreasing" everywhere, the word "smallest" is replaced by "largest", and the word "lowest" is replaced by "highest".

(c) That said, the decreasing list and the increasing list are closely related: If S is a finite set of integers (or rational numbers, or real numbers), and if $(s_1, s_2, \ldots, s_k)$ is the increasing list of S, then $(s_k, s_{k-1}, \ldots, s_1)$ is the decreasing list of S. (The proof is very simple.)

(d) Let S be a nonempty finite set of integers (or rational numbers, or real numbers), and let $(s_1, s_2, \ldots, s_k)$ be the increasing list of S. Proposition 2.49 **(a)** (applied to $m = \max S$) shows that $k \geq 1$ and $s_k = \max S$. A similar argument can be used to show that $s_1 = \min S$. Thus, the increasing list of S begins with the smallest element of S and ends with the largest element of S (as one would expect).

2.7. Induction with shifted base

2.7.1. Induction starting at g

All the induction proofs we have done so far were applications of Theorem 2.1 (even though we have often written them up in ways that hide the exact statements $\mathcal{A}(n)$ to which Theorem 2.1 is being applied). We are soon going to see several other "induction principles" which can also be used to make proofs. Unlike Theorem 2.1, these other principles need not be taken on trust; instead, they can themselves be proven using Theorem 2.1. Thus, they merely offer convenience, not new logical opportunities.

Our first such "alternative induction principle" is Theorem 2.53 below. First, we introduce a simple notation:

Definition 2.52. Let $g \in \mathbb{Z}$. Then, $\mathbb{Z}_{\geq g}$ denotes the set $\{g, g+1, g+2, \ldots\}$; this is the set of all integers that are $\geq g$.

For example, $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \ldots\} = \mathbb{N}$ is the set of all nonnegative integers, whereas $\mathbb{Z}_{\geq 1} = \{1, 2, 3, \ldots\}$ is the set of all positive integers.

Now, we state our first "alternative induction principle":

Theorem 2.53. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: The statement $\mathcal{A}(g)$ holds.

Assumption 2: If $m \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Again, Theorem 2.53 is intuitively clear: For example, if you have $g = 4$, and you want to prove (under the assumptions of Theorem 2.53) that $\mathcal{A}(8)$ holds, you can argue as follows:

- By Assumption 1, the statement $\mathcal{A}(4)$ holds.
- Thus, by Assumption 2 (applied to $m = 4$), the statement $\mathcal{A}(5)$ holds.
- Thus, by Assumption 2 (applied to $m = 5$), the statement $\mathcal{A}(6)$ holds.
- Thus, by Assumption 2 (applied to $m = 6$), the statement $\mathcal{A}(7)$ holds.
- Thus, by Assumption 2 (applied to $m = 7$), the statement $\mathcal{A}(8)$ holds.

A similar (but longer) argument shows that the statement $\mathcal{A}(9)$ holds; likewise, $\mathcal{A}(n)$ can be shown to hold for each $n \in \mathbb{Z}_{\geq g}$ by means of an argument that takes $n - g + 1$ steps.

Theorem 2.53 generalizes Theorem 2.1. Indeed, Theorem 2.1 is the particular case of Theorem 2.53 for $g = 0$ (since $\mathbb{Z}_{\geq 0} = \mathbb{N}$). However, Theorem 2.53 can also be derived from Theorem 2.1. In order to do this, we essentially need to "shift" the index n in Theorem 2.53 down by g – that is, we need to rename our sequence $(\mathcal{A}(g), \mathcal{A}(g+1), \mathcal{A}(g+2), \ldots)$ of statements as $(\mathcal{B}(0), \mathcal{B}(1), \mathcal{B}(2), \ldots)$, and apply Theorem 2.1 to $\mathcal{B}(n)$ instead of $\mathcal{A}(n)$. In order to make this renaming procedure rigorous, let us first restate Theorem 2.1 as follows:

Corollary 2.54. For each $n \in \mathbb{N}$, let $\mathcal{B}(n)$ be a logical statement.

Assume the following:

Assumption A: The statement $\mathcal{B}(0)$ holds.

Assumption B: If $p \in \mathbb{N}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds.

Then, $\mathcal{B}(n)$ holds for each $n \in \mathbb{N}$.

Proof of Corollary 2.54. Corollary 2.54 is exactly Theorem 2.1, except that some names have been changed:

- The statements $\mathcal{A}(n)$ have been renamed as $\mathcal{B}(n)$.
- Assumption 1 and Assumption 2 have been renamed as Assumption A and Assumption B.
- The variable m in Assumption B has been renamed as p.

Thus, Corollary 2.54 holds (since Theorem 2.1 holds). □

Let us now derive Theorem 2.53 from Theorem 2.1:

Proof of Theorem 2.53. For any $n \in \mathbb{N}$, we have $n + g \in \mathbb{Z}_{\geq g}$ [60]. Hence, for each $n \in \mathbb{N}$, we can define a logical statement $\mathcal{B}(n)$ by

$$\mathcal{B}(n) = \mathcal{A}(n+g).$$

Consider this $\mathcal{B}(n)$.

Now, let us consider the Assumptions A and B from Corollary 2.54. We claim that both of these assumptions are satisfied.

Indeed, the statement $\mathcal{A}(g)$ holds (by Assumption 1). But the definition of the statement $\mathcal{B}(0)$ shows that $\mathcal{B}(0) = \mathcal{A}(0+g) = \mathcal{A}(g)$. Hence, the statement $\mathcal{B}(0)$ holds (since the statement $\mathcal{A}(g)$ holds). In other words, Assumption A is satisfied.

Now, we shall show that Assumption B is satisfied. Indeed, let $p \in \mathbb{N}$ be such that $\mathcal{B}(p)$ holds. The definition of the statement $\mathcal{B}(p)$ shows that $\mathcal{B}(p) = \mathcal{A}(p+g)$. Hence, the statement $\mathcal{A}(p+g)$ holds (since $\mathcal{B}(p)$ holds).

Also, $p \in \mathbb{N}$, so that $p \geq 0$ and thus $p + g \geq g$. In other words, $p + g \in \mathbb{Z}_{\geq g}$ (since $\mathbb{Z}_{\geq g}$ is the set of all integers that are $\geq g$).

Recall that Assumption 2 holds. In other words, if $m \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. Applying this to $m = p + g$, we conclude that $\mathcal{A}((p+g)+1)$ holds (since $\mathcal{A}(p+g)$ holds).

But the definition of $\mathcal{B}(p+1)$ yields $\mathcal{B}(p+1) = \mathcal{A}\left(\underbrace{p+1+g}_{=(p+g)+1}\right) = \mathcal{A}((p+g)+1)$. Hence, the statement $\mathcal{B}(p+1)$ holds (since the statement $\mathcal{A}((p+g)+1)$ holds).

Now, forget that we fixed p. We thus have shown that if $p \in \mathbb{N}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds. In other words, Assumption B is satisfied.

We now know that both Assumption A and Assumption B are satisfied. Hence, Corollary 2.54 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{N}. \tag{83}$$

Now, let $n \in \mathbb{Z}_{\geq g}$. Thus, n is an integer such that $n \geq g$ (by the definition of $\mathbb{Z}_{\geq g}$). Hence, $n - g \geq 0$, so that $n - g \in \mathbb{N}$. Thus, (83) (applied to $n - g$

[60] *Proof.* Let $n \in \mathbb{N}$. Thus, $n \geq 0$, so that $\underbrace{n}_{\geq 0} + g \geq 0 + g = g$. Hence, $n + g$ is an integer $\geq g$. In other words, $n + g \in \mathbb{Z}_{\geq g}$ (since $\mathbb{Z}_{\geq g}$ is the set of all integers that are $\geq g$). Qed.

instead of n) yields that $\mathcal{B}(n-g)$ holds. But the definition of $\mathcal{B}(n-g)$ yields $\mathcal{B}(n-g) = \mathcal{A}\left(\underbrace{(n-g)+g}_{=n}\right) = \mathcal{A}(n)$. Hence, the statement $\mathcal{A}(n)$ holds (since $\mathcal{B}(n-g)$ holds).

Now, forget that we fixed n. We thus have shown that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$. This proves Theorem 2.53. □

Theorem 2.53 is called the *principle of induction starting at* g, and proofs that use it are usually called *proofs by induction* or *induction proofs*. As with the standard induction principle (Theorem 2.1), we don't usually explicitly cite Theorem 2.53, but instead say certain words that signal that it is being applied and that (ideally) also indicate what integer g and what statements $\mathcal{A}(n)$ it is being applied to[61]. However, for our very first example of the use of Theorem 2.53, we are going to reference it explicitly:

Proposition 2.55. Let a and b be integers. Then, every positive integer n satisfies

$$(a+b)^n \equiv a^n + na^{n-1}b \mod b^2. \tag{84}$$

Note that we have chosen not to allow $n = 0$ in Proposition 2.55, because it is not clear what "a^{n-1}" would mean when $n = 0$ and $a = 0$. (Recall that $0^{0-1} = 0^{-1}$ is not defined!) In truth, it is easy to convince oneself that this is not a serious hindrance, since the expression "na^{n-1}" has a meaningful interpretation even when its subexpression "a^{n-1}" does not (one just has to interpret it as 0 when $n = 0$, without regard to whether "a^{n-1}" is well-defined). Nevertheless, we prefer to rule out the case of $n = 0$ by requiring n to be positive, in order to avoid having to discuss such questions of interpretation. (Of course, this also gives us an excuse to apply Theorem 2.53 instead of the old Theorem 2.1.)

Proof of Proposition 2.55. For each $n \in \mathbb{Z}_{\geq 1}$, we let $\mathcal{A}(n)$ be the statement

$$\left((a+b)^n \equiv a^n + na^{n-1}b \mod b^2\right).$$

Our next goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \mathbb{Z}_{\geq 1}$.

We first notice that the statement $\mathcal{A}(1)$ holds[62].

Now, we claim that

$$\text{if } m \in \mathbb{Z}_{\geq 1} \text{ is such that } \mathcal{A}(m) \text{ holds, then } \mathcal{A}(m+1) \text{ also holds.} \tag{85}$$

[61] We will explain this in Convention 2.56 below.

[62] *Proof.* We have $(a+b)^1 = a+b$. Comparing this with $\underbrace{a^1}_{=a} + 1 \underbrace{a^{1-1}}_{=a^0=1} b = a+b$, we obtain $(a+b)^1 = a^1 + 1a^{1-1}b$. Hence, $(a+b)^1 \equiv a^1 + 1a^{1-1}b \mod b^2$. But this is precisely the statement $\mathcal{A}(1)$ (since $\mathcal{A}(1)$ is defined to be the statement $\left((a+b)^1 \equiv a^1 + 1a^{1-1}b \mod b^2\right)$). Hence, the statement $\mathcal{A}(1)$ holds.

[*Proof of (85):* Let $m \in \mathbb{Z}_{\geq 1}$ be such that $\mathcal{A}(m)$ holds. We must show that $\mathcal{A}(m+1)$ also holds.

We have assumed that $\mathcal{A}(m)$ holds. In other words,

$$(a+b)^m \equiv a^m + ma^{m-1}b \operatorname{mod} b^2$$

holds[63]. Now,

$$\begin{aligned}
(a+b)^{m+1} &= \underbrace{(a+b)^m}_{\equiv a^m + ma^{m-1}b \operatorname{mod} b^2} (a+b) \\
&\equiv \left(a^m + ma^{m-1}b\right)(a+b) \\
&= \underbrace{a^m a}_{=a^{m+1}} + a^m b + m \underbrace{a^{m-1}ba}_{\substack{=a^{m-1}ab = a^m b \\ (\text{since } a^{m-1}a = a^m)}} + \underbrace{ma^{m-1}bb}_{\substack{=ma^{m-1}b^2 \equiv 0 \operatorname{mod} b^2 \\ (\text{since } b^2 \mid ma^{m-1}b^2)}} \\
&\equiv a^{m+1} + \underbrace{a^m b + ma^m b}_{=(m+1)a^m b} + 0 \\
&= a^{m+1} + (m+1) \underbrace{a^m}_{\substack{=a^{(m+1)-1} \\ (\text{since } m=(m+1)-1)}} b = a^{m+1} + (m+1)\, a^{(m+1)-1} b \operatorname{mod} b^2.
\end{aligned}$$

So we have shown that $(a+b)^{m+1} \equiv a^{m+1} + (m+1)\, a^{(m+1)-1} b \operatorname{mod} b^2$. But this is precisely the statement $\mathcal{A}(m+1)$ [64]. Thus, the statement $\mathcal{A}(m+1)$ holds.

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{Z}_{\geq 1}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. This proves (85).]

Now, both assumptions of Theorem 2.53 (applied to $g = 1$) are satisfied (indeed, Assumption 1 is satisfied because the statement $\mathcal{A}(1)$ holds, whereas Assumption 2 is satisfied because of (85)). Thus, Theorem 2.53 (applied to $g = 1$) shows that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq 1}$. In other words, $(a+b)^n \equiv a^n + na^{n-1}b \operatorname{mod} b^2$ holds for each $n \in \mathbb{Z}_{\geq 1}$ (since $\mathcal{A}(n)$ is the statement $\left((a+b)^n \equiv a^n + na^{n-1}b \operatorname{mod} b^2\right)$). In other words, $(a+b)^n \equiv a^n + na^{n-1}b \operatorname{mod} b^2$ holds for each positive integer n (because the positive integers are exactly the $n \in \mathbb{Z}_{\geq 1}$). This proves Proposition 2.55. □

2.7.2. Conventions for writing proofs by induction starting at g

Now, let us introduce some standard language that is commonly used in proofs by induction starting at g:

[63] because $\mathcal{A}(m)$ is defined to be the statement $\left((a+b)^m \equiv a^m + ma^{m-1}b \operatorname{mod} b^2\right)$

[64] because $\mathcal{A}(m+1)$ is defined to be the statement $\left((a+b)^{m+1} \equiv a^{m+1} + (m+1)\, a^{(m+1)-1} b \operatorname{mod} b^2\right)$

Convention 2.56. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{A}(n)$ be a logical statement. Assume that you want to prove that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Theorem 2.53 offers the following strategy for proving this: First show that Assumption 1 of Theorem 2.53 is satisfied; then, show that Assumption 2 of Theorem 2.53 is satisfied; then, Theorem 2.53 automatically completes your proof.

A proof that follows this strategy is called a *proof by induction on n* (or *proof by induction over n*) *starting at g* or (less precisely) an *inductive proof*. Most of the time, the words "starting at g" are omitted, since they merely repeat what is clear from the context anyway: For example, if you make a claim about all integers $n \geq 3$, and you say that you are proving it by induction on n, then it is clear that you are using induction on n starting at 3. (And if this isn't clear from the claim, then the induction base will make it clear.)

The proof that Assumption 1 is satisfied is called the *induction base* (or *base case*) of the proof. The proof that Assumption 2 is satisfied is called the *induction step* of the proof.

In order to prove that Assumption 2 is satisfied, you will usually want to fix an $m \in \mathbb{Z}_{\geq g}$ such that $\mathcal{A}(m)$ holds, and then prove that $\mathcal{A}(m+1)$ holds. In other words, you will usually want to fix $m \in \mathbb{Z}_{\geq g}$, assume that $\mathcal{A}(m)$ holds, and then prove that $\mathcal{A}(m+1)$ holds. When doing so, it is common to refer to the assumption that $\mathcal{A}(m)$ holds as the *induction hypothesis* (or *induction assumption*).

Unsurprisingly, this language parallels the language introduced in Convention 2.3 for proofs by "standard" induction.

Again, we can shorten our inductive proofs by omitting some sentences that convey no information. In particular, we can leave out the explicit definition of the statement $\mathcal{A}(n)$ when this statement is precisely the claim that we are proving (without the "for each $n \in \mathbb{Z}_{\geq g}$" part). Thus, we can rewrite our above proof of Proposition 2.55 as follows:

Proof of Proposition 2.55 (second version). We must prove (84) for every positive integer n. In other words, we must prove (84) for every $n \in \mathbb{Z}_{\geq 1}$ (since the positive integers are precisely the $n \in \mathbb{Z}_{\geq 1}$). We shall prove this by induction on n starting at 1:

Induction base: We have $(a+b)^1 = a+b$. Comparing this with $\underbrace{a^1}_{=a} + 1 \underbrace{a^{1-1}}_{=a^0=1} b = a+b$, we obtain $(a+b)^1 = a^1 + 1a^{1-1}b$. Hence, $(a+b)^1 \equiv a^1 + 1a^{1-1}b \mod b^2$. In other words, (84) holds for $n=1$. This completes the induction base.

Induction step: Let $m \in \mathbb{Z}_{\geq 1}$. Assume that (84) holds for $n = m$. We must show that (84) also holds for $n = m+1$.

We have assumed that (84) holds for $n = m$. In other words,

$$(a+b)^m \equiv a^m + ma^{m-1}b \mod b^2$$

holds. Now,

$$\begin{aligned}
(a+b)^{m+1} &= \underbrace{(a+b)^m}_{\equiv a^m + ma^{m-1}b \operatorname{mod} b^2} (a+b) \\
&\equiv \left(a^m + ma^{m-1}b\right)(a+b) \\
&= \underbrace{a^m a}_{=a^{m+1}} + a^m b + m \underbrace{a^{m-1}ba}_{\substack{=a^{m-1}ab = a^m b \\ (\text{since } a^{m-1}a = a^m)}} + \underbrace{ma^{m-1}bb}_{\substack{=ma^{m-1}b^2 \equiv 0 \operatorname{mod} b^2 \\ (\text{since } b^2 \mid ma^{m-1}b^2)}} \\
&\equiv a^{m+1} + \underbrace{a^m b + ma^m b}_{=(m+1)a^m b} + 0 \\
&= a^{m+1} + (m+1) \underbrace{a^m}_{\substack{=a^{(m+1)-1} \\ (\text{since } m = (m+1)-1)}} b = a^{m+1} + (m+1)\, a^{(m+1)-1} b \operatorname{mod} b^2.
\end{aligned}$$

So we have shown that $(a+b)^{m+1} \equiv a^{m+1} + (m+1)\, a^{(m+1)-1} b \operatorname{mod} b^2$. In other words, (84) holds for $n = m+1$.

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{Z}_{\geq 1}$ is such that (84) holds for $n = m$, then (84) also holds for $n = m+1$. This completes the induction step. Hence, (84) is proven by induction. This proves Proposition 2.55. □

Proposition 2.55 can also be seen as a consequence of the binomial formula (Proposition 3.21 further below).

2.7.3. More properties of congruences

Let us use this occasion to show two corollaries of Proposition 2.55:

Corollary 2.57. Let a, b and n be three integers such that $a \equiv b \operatorname{mod} n$. Let $d \in \mathbb{N}$ be such that $d \mid n$. Then, $a^d \equiv b^d \operatorname{mod} nd$.

Proof of Corollary 2.57. We have $a \equiv b \operatorname{mod} n$. In other words, a is congruent to b modulo n. In other words, $n \mid a - b$ (by the definition of "congruent"). In other words, there exists an integer w such that $a - b = nw$. Consider this w. From $a - b = nw$, we obtain $a = b + nw$. Also, $d \mid n$, thus $dn \mid nn$ (by Proposition 2.6, applied to d, n and n instead of a, b and c). On the other hand, $nn \mid (nw)^2$ (since $(nw)^2 = nwnw = nnww$). Hence, Proposition 2.5 (applied to dn, nn and $(nw)^2$ instead of a, b and c) yields $dn \mid (nw)^2$ (since $dn \mid nn$ and $nn \mid (nw)^2$). In other words, $nd \mid (nw)^2$ (since $dn = nd$).

Next, we claim that

$$nd \mid a^d - b^d. \tag{86}$$

[*Proof of (86):* If $d = 0$, then (86) holds (because if $d = 0$, then $a^d - b^d = \underbrace{a^0}_{=1} - \underbrace{b^0}_{=1} = 1 - 1 = 0 = 0nd$, and thus $nd \mid a^d - b^d$). Hence, for the rest of this proof of (86), we WLOG assume that we don't have $d = 0$. Thus, $d \neq 0$. Hence, d is a positive integer (since $d \in \mathbb{N}$). Thus, Proposition 2.55 (applied to d, b and nw instead of n, a and b) yields

$$(b + nw)^d \equiv b^d + db^{d-1} nw \mod (nw)^2 .$$

In view of $a = b + nw$, this rewrites as

$$a^d \equiv b^d + db^{d-1} nw \mod (nw)^2 .$$

Hence, Proposition 2.11 **(c)** (applied to a^d, $b^d + db^{d-1}nw$, $(nw)^2$ and nd instead of a, b, n and m) yields

$$a^d \equiv b^d + db^{d-1} nw \mod nd$$

(since $nd \mid (nw)^2$). Hence,

$$a^d \equiv b^d + \underbrace{db^{d-1}nw}_{\substack{= ndb^{d-1}w \equiv 0 \mod nd \\ (\text{since } nd \mid ndb^{d-1}w)}} \equiv b^d + 0 = b^d \mod nd.$$

In other words, $nd \mid a^d - b^d$. This proves (86).]

From (86), we immediately obtain $a^d \equiv b^d \mod nd$ (by the definition of "congruent"). This proves Corollary 2.57. □

For the next corollary, we need a convention:

Convention 2.58. Let a, b and c be three integers. Then, the expression "a^{b^c}" shall always be interpreted as "$a^{(b^c)}$", never as "$\left(a^b\right)^c$".

Thus, for example, "3^{3^3}" means $3^{(3^3)} = 3^{27} = 7625\,597\,484\,987$, not $\left(3^3\right)^3 = 27^3 = 19\,683$. The reason for this convention is that $\left(a^b\right)^c$ can be simplified to a^{bc} and thus there is little use in having yet another notation for it. Of course, this convention applies not only to integers, but to any other numbers a, b, c.

We can now state the following fact, which is sometimes known as "lifting-the-exponent lemma":

Corollary 2.59. Let $n \in \mathbb{N}$. Let a and b be two integers such that $a \equiv b \mod n$. Let $k \in \mathbb{N}$. Then,

$$a^{n^k} \equiv b^{n^k} \mod n^{k+1}. \tag{87}$$

We shall give two **different** proofs of Corollary 2.59 by induction on k, to illustrate once again the point (previously made in Remark 2.27) that we have a choice of what precise statement we are proving by induction. In the first proof, the statement will be the congruence (87) for three **fixed** integers a, b and n, whereas in the second proof, it will be the statement

$$\left(a^{n^k} \equiv b^{n^k} \mod n^{k+1} \text{ for **all** integers } a \text{ and } b \text{ and **all** } n \in \mathbb{N} \text{ satisfying } a \equiv b \mod n\right).$$

First proof of Corollary 2.59. Forget that we fixed k. We thus must prove (87) for each $k \in \mathbb{N}$.

We shall prove this by induction on k:

Induction base: We have $n^0 = 1$ and thus $a^{n^0} = a^1 = a$. Similarly, $b^{n^0} = b$. Thus, $a^{n^0} = a \equiv b = b^{n^0} \operatorname{mod} n$. In other words, $a^{n^0} \equiv b^{n^0} \operatorname{mod} n^{0+1}$ (since $n^{0+1} = n^1 = n$). In other words, (87) holds for $k = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (87) holds for $k = m$. We must prove that (87) holds for $k = m + 1$.

We have $n^{m+1} = nn^m$. Hence, $n \mid n^{m+1}$.

We have assumed that (87) holds for $k = m$. In other words, we have

$$a^{n^m} \equiv b^{n^m} \operatorname{mod} n^{m+1}.$$

Hence, Corollary 2.57 (applied to a^{n^m}, b^{n^m}, n^{m+1} and n instead of a, b, n and d) yields

$$\left(a^{n^m}\right)^n \equiv \left(b^{n^m}\right)^n \operatorname{mod} n^{m+1}n.$$

Now, $n^{m+1} = n^m n$, so that

$$a^{n^{m+1}} = a^{n^m n} = \left(a^{n^m}\right)^n \equiv \left(b^{n^m}\right)^n = b^{n^m n} = b^{n^{m+1}} \operatorname{mod} n^{m+1}n$$

(since $n^m n = n^{m+1}$). In view of $n^{m+1}n = n^{(m+1)+1}$, this rewrites as

$$a^{n^{m+1}} \equiv b^{n^{m+1}} \operatorname{mod} n^{(m+1)+1}.$$

In other words, (87) holds for $k = m + 1$. This completes the induction step. Thus, (87) is proven by induction. Hence, Corollary 2.59 holds. □

Second proof of Corollary 2.59. Forget that we fixed a, b, n and k. We thus must prove

$$\left(a^{n^k} \equiv b^{n^k} \operatorname{mod} n^{k+1} \text{ for all integers } a \text{ and } b \text{ and all } n \in \mathbb{N} \text{ satisfying } a \equiv b \operatorname{mod} n\right) \tag{88}$$

for all $k \in \mathbb{N}$.

We shall prove this by induction on k:

Induction base: Let $n \in \mathbb{N}$. Let a and b be two integers such that $a \equiv b \operatorname{mod} n$. We have $n^0 = 1$ and thus $a^{n^0} = a^1 = a$. Similarly, $b^{n^0} = b$. Thus, $a^{n^0} = a \equiv b = b^{n^0} \operatorname{mod} n$. In other words, $a^{n^0} \equiv b^{n^0} \operatorname{mod} n^{0+1}$ (since $n^{0+1} = n^1 = n$).

Now, forget that we fixed n, a and b. We thus have proven that $a^{n^0} \equiv b^{n^0} \operatorname{mod} n^{0+1}$ for all integers a and b and all $n \in \mathbb{N}$ satisfying $a \equiv b \operatorname{mod} n$. In other words, (88) holds for $k = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (88) holds for $k = m$. We must prove that (88) holds for $k = m + 1$.

Let $n \in \mathbb{N}$. Let a and b be two integers such that $a \equiv b \mod n$. Now,

$$\left(n^2\right)^{m+1} = n^{2(m+1)} = n^{(m+2)+m} \qquad \left(\text{since } 2\left(m+1\right) = \left(m+2\right)+m\right)$$
$$= n^{m+2}n^m,$$

so that $n^{m+2} \mid \left(n^2\right)^{m+1}$.

We have $n \mid n$. Hence, Corollary 2.57 (applied to $d = n$) yields $a^n \equiv b^n \mod nn$. In other words, $a^n \equiv b^n \mod n^2$ (since $nn = n^2$).

We have assumed that (88) holds for $k = m$. Hence, we can apply (88) to a^n, b^n, n^2 and m instead of a, b, n and k (since $a^n \equiv b^n \mod n^2$). We thus conclude that

$$\left(a^n\right)^{n^m} \equiv \left(b^n\right)^{n^m} \mod \left(n^2\right)^{m+1}.$$

Now, $n^{m+1} = nn^m$, so that

$$a^{n^{m+1}} = a^{nn^m} = \left(a^n\right)^{n^m} \equiv \left(b^n\right)^{n^m} = b^{nn^m} = b^{n^{m+1}} \mod \left(n^2\right)^{m+1}$$

(since $nn^m = n^{m+1}$). Hence, Proposition 2.11 **(c)** (applied to $a^{n^{m+1}}$, $b^{n^{m+1}}$, $\left(n^2\right)^{m+1}$ and n^{m+2} instead of a, b, n and m) yields $a^{n^{m+1}} \equiv b^{n^{m+1}} \mod n^{m+2}$ (since $n^{m+2} \mid \left(n^2\right)^{m+1}$). In view of $m+2 = (m+1)+1$, this rewrites as

$$a^{n^{m+1}} \equiv b^{n^{m+1}} \mod n^{(m+1)+1}.$$

Now, forget that we fixed n, a and b. We thus have proven that $a^{n^{m+1}} \equiv b^{n^{m+1}} \mod n^{(m+1)+1}$ for all integers a and b and all $n \in \mathbb{N}$ satisfying $a \equiv b \mod n$. In other words, (88) holds for $k = m+1$. This completes the induction step. Thus, (88) is proven by induction. Hence, Corollary 2.59 is proven again. □

2.8. Strong induction

2.8.1. The strong induction principle

We shall now show another "alternative induction principle", which is known as the *strong induction principle* because it feels stronger than Theorem 2.1 (in the sense that it appears to get the same conclusion from weaker assumptions). Just as Theorem 2.53, this principle is not a new axiom, but rather a consequence of the standard induction principle; we shall soon deduce it from Theorem 2.53.

Theorem 2.60. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: If $m \in \mathbb{Z}_{\geq g}$ is such that

$$\left(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < m\right),$$

then $\mathcal{A}(m)$ holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Notice that Theorem 2.60 has only one assumption (unlike Theorem 2.1 and Theorem 2.53). We shall soon see that this one assumption "incorporates" both an induction base and an induction step.

Let us first explain why Theorem 2.60 is intuitively clear. For example, if you have $g = 4$, and you want to prove (under the assumptions of Theorem 2.60) that $\mathcal{A}(7)$ holds, you can argue as follows:

- We know that $\mathcal{A}(n)$ holds for every $n \in \mathbb{Z}_{\geq 4}$ satisfying $n < 4$. (Indeed, this is vacuously true, since there is no $n \in \mathbb{Z}_{\geq 4}$ satisfying $n < 4$.)

  Hence, Assumption 1 (applied to $m = 4$) shows that the statement $\mathcal{A}(4)$ holds.

- Thus, we know that $\mathcal{A}(n)$ holds for every $n \in \mathbb{Z}_{\geq 4}$ satisfying $n < 5$ (because $\mathcal{A}(4)$ holds).

  Hence, Assumption 1 (applied to $m = 5$) shows that the statement $\mathcal{A}(5)$ holds.

- Thus, we know that $\mathcal{A}(n)$ holds for every $n \in \mathbb{Z}_{\geq 4}$ satisfying $n < 6$ (because $\mathcal{A}(4)$ and $\mathcal{A}(5)$ hold).

  Hence, Assumption 1 (applied to $m = 6$) shows that the statement $\mathcal{A}(6)$ holds.

- Thus, we know that $\mathcal{A}(n)$ holds for every $n \in \mathbb{Z}_{\geq 4}$ satisfying $n < 7$ (because $\mathcal{A}(4)$, $\mathcal{A}(5)$ and $\mathcal{A}(6)$ hold).

  Hence, Assumption 1 (applied to $m = 7$) shows that the statement $\mathcal{A}(7)$ holds.

A similar (but longer) argument shows that the statement $\mathcal{A}(8)$ holds; likewise, $\mathcal{A}(n)$ can be shown to hold for each $n \in \mathbb{Z}_{\geq g}$ by means of an argument that takes $n - g + 1$ steps.

It is easy to see that Theorem 2.60 generalizes Theorem 2.53 (because if the two Assumptions 1 and 2 of Theorem 2.53 hold, then so does Assumption 1 of Theorem 2.60). More interesting for us is the converse implication: We shall show that Theorem 2.60 can be derived from Theorem 2.53. This will allow us to use Theorem 2.60 without having to taking it on trust.

Before we derive Theorem 2.60, let us restate Theorem 2.53 as follows:

Corollary 2.61. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{B}(n)$ be a logical statement.

Assume the following:

Assumption A: The statement $\mathcal{B}(g)$ holds.

Assumption B: If $p \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds.

Then, $\mathcal{B}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Proof of Corollary 2.61. Corollary 2.61 is exactly Theorem 2.53, except that some names have been changed:

- The statements $\mathcal{A}(n)$ have been renamed as $\mathcal{B}(n)$.
- Assumption 1 and Assumption 2 have been renamed as Assumption A and Assumption B.
- The variable m in Assumption B has been renamed as p.

Thus, Corollary 2.61 holds (since Theorem 2.53 holds). □

Let us now derive Theorem 2.60 from Theorem 2.53:

Proof of Theorem 2.60. For each $n \in \mathbb{Z}_{\geq g}$, we let $\mathcal{B}(n)$ be the statement

$$\left(\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < n\right).$$

Now, let us consider the Assumptions A and B from Corollary 2.61. We claim that both of these assumptions are satisfied.

The statement $\mathcal{B}(g)$ holds[65]. Thus, Assumption A is satisfied.

Next, let us prove that Assumption B is satisfied. Indeed, let $p \in \mathbb{Z}_{\geq g}$ be such that $\mathcal{B}(p)$ holds. We shall show that $\mathcal{B}(p+1)$ also holds.

Indeed, we have assumed that $\mathcal{B}(p)$ holds. In other words,

$$\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < p \tag{89}$$

(because the statement $\mathcal{B}(p)$ is defined as $\left(\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < p\right)$). Renaming the variable q as n in this statement, we conclude that

$$\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < p. \tag{90}$$

Hence, Assumption 1 (applied to $m = p$) yields that $\mathcal{A}(p)$ holds.

Now, we claim that

$$\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < p + 1. \tag{91}$$

[*Proof of (91):* Let $q \in \mathbb{Z}_{\geq g}$ be such that $q < p + 1$. We must prove that $\mathcal{A}(q)$ holds.

[65] *Proof.* Let $q \in \mathbb{Z}_{\geq g}$ be such that $q < g$. Then, $q \geq g$ (since $q \in \mathbb{Z}_{\geq g}$); but this contradicts $q < g$.

Now, forget that we fixed q. We thus have found a contradiction for each $q \in \mathbb{Z}_{\geq g}$ satisfying $q < g$. Hence, there exists no $q \in \mathbb{Z}_{\geq g}$ satisfying $q < g$. Thus, the statement

$$\left(\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < g\right)$$

is vacuously true, and therefore true. In other words, the statement $\mathcal{B}(g)$ is true (since $\mathcal{B}(g)$ is defined as the statement $\left(\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < g\right)$). Qed.

If $q = p$, then this follows from the fact that $\mathcal{A}(p)$ holds. Hence, for the rest of this proof, we WLOG assume that we don't have $q = p$. Thus, $q \neq p$. But $q < p+1$ and therefore $q \leq (p+1) - 1$ (since q and $p+1$ are integers). Hence, $q \leq (p+1) - 1 = p$. Combining this with $q \neq p$, we obtain $q < p$. Hence, (89) shows that $\mathcal{A}(q)$ holds. This completes the proof of (91).]

But the statement $\mathcal{B}(p+1)$ is defined as $\left(\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < p+1\right)$. In other words, the statement $\mathcal{B}(p+1)$ is precisely the statement (91). Hence, the statement $\mathcal{B}(p+1)$ holds (since (91) holds).

Now, forget that we fixed p. We thus have shown that if $p \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds. In other words, Assumption B is satisfied.

We now know that both Assumption A and Assumption B are satisfied. Hence, Corollary 2.61 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{Z}_{\geq g}. \tag{92}$$

Now, let $n \in \mathbb{Z}_{\geq g}$. Thus, n is an integer such that $n \geq g$ (by the definition of $\mathbb{Z}_{\geq g}$). Hence, $n+1$ is also an integer and satisfies $n+1 \geq n \geq g$, so that $n+1 \in \mathbb{Z}_{\geq g}$. Hence, (92) (applied to $n+1$ instead of n) shows that $\mathcal{B}(n+1)$ holds. In other words,

$$\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < n+1$$

(because the statement $\mathcal{B}(n+1)$ is defined as $\left(\mathcal{A}(q) \text{ holds for every } q \in \mathbb{Z}_{\geq g} \text{ satisfying } q < n+1\right)$). We can apply this to $q = n$ (because $n \in \mathbb{Z}_{\geq g}$ satisfies $n < n+1$), and conclude that $\mathcal{A}(n)$ holds.

Now, forget that we fixed n. We thus have shown that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$. This proves Theorem 2.60. □

Thus, proving a sequence of statements $\mathcal{A}(0), \mathcal{A}(1), \mathcal{A}(2), \ldots$ using Theorem 2.60 is tantamount to proving a slightly different sequence of statements $\mathcal{B}(0), \mathcal{B}(1), \mathcal{B}(2), \ldots$ using Corollary 2.61 and then deriving the former from the latter.

Theorem 2.53 is called the *principle of strong induction starting at g*, and proofs that use it are usually called *proofs by strong induction*. We illustrate its use on the following easy property of the Fibonacci sequence:

Proposition 2.62. Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence (defined as in Example 2.25). Then,

$$f_n \leq 2^{n-1} \tag{93}$$

for each $n \in \mathbb{N}$.

Proof of Proposition 2.62. For each $n \in \mathbb{Z}_{\geq 0}$, we let $\mathcal{A}(n)$ be the statement $\left(f_n \leq 2^{n-1}\right)$.

Thus, $\mathcal{A}(0)$ is the statement $\left(f_0 \leq 2^{0-1}\right)$; hence, this statement holds (since $f_0 = 0 \leq 2^{0-1}$).

Also, $\mathcal{A}(1)$ is the statement $\left(f_1 \leq 2^{1-1}\right)$ (by the definition of $\mathcal{A}(1)$); hence, this statement also holds (since $f_1 = 1 = 2^{1-1}$).

Now, we claim the following:

Claim 1: If $m \in \mathbb{Z}_{\geq 0}$ is such that

$$(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq 0} \text{ satisfying } n < m),$$

then $\mathcal{A}(m)$ holds.

[*Proof of Claim 1:* Let $m \in \mathbb{Z}_{\geq 0}$ be such that

$$(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq 0} \text{ satisfying } n < m). \tag{94}$$

We must prove that $\mathcal{A}(m)$ holds.

This is true if $m \in \{0, 1\}$ (because we have shown that both statements $\mathcal{A}(0)$ and $\mathcal{A}(1)$ hold). Thus, for the rest of the proof of Claim 1, we WLOG assume that we don't have $m \in \{0, 1\}$. Hence, $m \in \mathbb{N} \setminus \{0, 1\} = \{2, 3, 4, \ldots\}$, so that $m \geq 2$.

From $m \geq 2$, we conclude that $m - 1 \geq 2 - 1 = 1 \geq 0$ and $m - 2 \geq 2 - 2 = 0$. Thus, both $m - 1$ and $m - 2$ belong to $\mathbb{N}$; therefore, f_{m-1} and f_{m-2} are well-defined.

We have $m - 1 \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ and $m - 1 < m$. Hence, (94) (applied to $n = m - 1$) yields that $\mathcal{A}(m-1)$ holds. In other words, $f_{m-1} \leq 2^{(m-1)-1}$ (because this is what the statement $\mathcal{A}(m-1)$ says).

We have $m - 2 \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ and $m - 2 < m$. Hence, (94) (applied to $n = m - 2$) yields that $\mathcal{A}(m-2)$ holds. In other words, $f_{m-2} \leq 2^{(m-2)-1}$ (because this is what the statement $\mathcal{A}(m-2)$ says).

We have $(m-1) - 1 = m - 2$ and thus $2^{(m-1)-1} = 2^{m-2} = 2 \cdot 2^{(m-2)-1} \geq 2^{(m-2)-1}$ (since $2 \cdot 2^{(m-2)-1} - 2^{(m-2)-1} = 2^{(m-2)-1} \geq 0$). Hence, $2^{(m-2)-1} \leq 2^{(m-1)-1}$.

But the recursive definition of the Fibonacci sequence yields $f_m = f_{m-1} + f_{m-2}$ (since $m \geq 2$). Hence,

$$f_m = \underbrace{f_{m-1}}_{\leq 2^{(m-1)-1}} + \underbrace{f_{m-2}}_{\leq 2^{(m-2)-1} \leq 2^{(m-1)-1}} \leq 2^{(m-1)-1} + 2^{(m-1)-1} = 2 \cdot 2^{(m-1)-1} = 2^{m-1}.$$

In other words, the statement $\mathcal{A}(m)$ holds (since the statement $\mathcal{A}(m)$ is defined to be $\left(f_m \leq 2^{m-1}\right)$). This completes the proof of Claim 1.]

Claim 1 shows that Assumption 1 of Theorem 2.60 (applied to $g = 0$) is satisfied. Hence, Theorem 2.60 (applied to $g = 0$) shows that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq 0}$. In other words, $f_n \leq 2^{n-1}$ holds for each $n \in \mathbb{Z}_{\geq 0}$ (since the statement $\mathcal{A}(n)$ is defined to be $\left(f_n \leq 2^{n-1}\right)$). In other words, $f_n \leq 2^{n-1}$ holds for each $n \in \mathbb{N}$ (since $\mathbb{Z}_{\geq 0} = \mathbb{N}$). This proves Proposition 2.62. □

2.8.2. Conventions for writing strong induction proofs

Again, when using the principle of strong induction, one commonly does not directly cite Theorem 2.60; instead one uses the following language:

Convention 2.63. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{A}(n)$ be a logical statement. Assume that you want to prove that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Theorem 2.60 offers the following strategy for proving this: Show that Assumption 1 of Theorem 2.60 is satisfied; then, Theorem 2.60 automatically completes your proof.

A proof that follows this strategy is called a *proof by strong induction on n starting at g*. The proof that Assumption 1 is satisfied is called the *induction step* of the proof. This kind of proof does not have an "induction base" (unlike proofs that use Theorem 2.1 or Theorem 2.53).[66]

In order to prove that Assumption 1 is satisfied, you will usually want to fix an $m \in \mathbb{Z}_{\geq g}$ such that

$$\left(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < m\right), \tag{95}$$

and then prove that $\mathcal{A}(m)$ holds. In other words, you will usually want to fix $m \in \mathbb{Z}_{\geq g}$, assume that (95) holds, and then prove that $\mathcal{A}(m)$ holds. When doing so, it is common to refer to the assumption that (95) holds as the *induction hypothesis* (or *induction assumption*).

Using this language, we can rewrite our above proof of Proposition 2.62 as follows:

Proof of Proposition 2.62 (second version). For each $n \in \mathbb{Z}_{\geq 0}$, we let $\mathcal{A}(n)$ be the statement $\left(f_n \leq 2^{n-1}\right)$. Thus, our goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \mathbb{N}$. In other words, our goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \mathbb{Z}_{\geq 0}$ (since $\mathbb{N} = \mathbb{Z}_{\geq 0}$).

We shall prove this by strong induction on n starting at 0:

Induction step: Let $m \in \mathbb{Z}_{\geq 0}$. Assume that

$$\left(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq 0} \text{ satisfying } n < m\right). \tag{96}$$

We must then show that $\mathcal{A}(m)$ holds. In other words, we must show that $f_m \leq 2^{m-1}$ holds (since the statement $\mathcal{A}(m)$ is defined as $\left(f_m \leq 2^{m-1}\right)$).

This is true if $m = 0$ (since $f_0 = 0 \leq 2^{0-1}$) and also true if $m = 1$ (since $f_1 = 1 = 2^{1-1}$ and thus $f_1 \leq 2^{1-1}$). In other words, this is true if $m \in \{0, 1\}$. Thus, for the rest of the induction step, we WLOG assume that we don't have $m \in \{0, 1\}$. Hence, $m \notin \{0, 1\}$, so that $m \in \mathbb{N} \setminus \{0, 1\} = \{2, 3, 4, \ldots\}$. Hence, $m \geq 2$.

From $m \geq 2$, we conclude that $m - 1 \geq 2 - 1 = 1 \geq 0$ and $m - 2 \geq 2 - 2 = 0$. Thus, both $m - 1$ and $m - 2$ belong to $\mathbb{N}$; therefore, f_{m-1} and f_{m-2} are well-defined.

We have $m - 1 \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ and $m - 1 < m$. Hence, (96) (applied to $n = m - 1$) yields that $\mathcal{A}(m-1)$ holds. In other words, $f_{m-1} \leq 2^{(m-1)-1}$ (because this is what the statement $\mathcal{A}(m-1)$ says).

[66] There is a version of strong induction which does include an induction base (or even several). But the version we are using does not.

We have $m-2 \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ and $m-2 < m$. Hence, (96) (applied to $n = m-2$) yields that $\mathcal{A}(m-2)$ holds. In other words, $f_{m-2} \leq 2^{(m-2)-1}$ (because this is what the statement $\mathcal{A}(m-2)$ says).

We have $(m-1)-1 = m-2$ and thus $2^{(m-1)-1} = 2^{m-2} = 2 \cdot 2^{(m-2)-1} \geq 2^{(m-2)-1}$ (since $2 \cdot 2^{(m-2)-1} - 2^{(m-2)-1} = 2^{(m-2)-1} \geq 0$). Hence, $2^{(m-2)-1} \leq 2^{(m-1)-1}$.

But the recursive definition of the Fibonacci sequence yields $f_m = f_{m-1} + f_{m-2}$ (since $m \geq 2$). Hence,

$$f_m = \underbrace{f_{m-1}}_{\leq 2^{(m-1)-1}} + \underbrace{f_{m-2}}_{\leq 2^{(m-2)-1} \leq 2^{(m-1)-1}} \leq 2^{(m-1)-1} + 2^{(m-1)-1} = 2 \cdot 2^{(m-1)-1} = 2^{m-1}.$$

In other words, the statement $\mathcal{A}(m)$ holds (since the statement $\mathcal{A}(m)$ is defined to be $\left(f_m \leq 2^{m-1}\right)$).

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{Z}_{\geq 0}$ is such that (96) holds, then $\mathcal{A}(m)$ holds. This completes the induction step. Hence, by strong induction, we conclude that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq 0}$. This completes our proof of Proposition 2.62. □

The proof that we just showed still has a lot of "boilerplate" text that conveys no information. For example, we have again explicitly defined the statement $\mathcal{A}(n)$, which is unnecessary: This statement is exactly what one would expect (namely, the claim that we are proving, without the "for each $n \in \mathbb{N}$" part). Thus, in our case, this statement is simply (93). Furthermore, we can remove the two sentences

> "Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{Z}_{\geq 0}$ is such that (96) holds, then $\mathcal{A}(m)$ holds.".

In fact, these sentences merely say that we have completed the induction step; but this is clear anyway when we say that the induction step is completed.

We said that we are proving our statement "by strong induction on n starting at 0". Again, we can omit the words "starting at 0" here, since this is the only option (because our statement is about all $n \in \mathbb{Z}_{\geq 0}$).

Finally, we can remove the words "*Induction step:*", because a proof by strong induction (unlike a proof by standard induction) does not have an induction base (so the induction step is all that it consists of).

Thus, our above proof can be shortened to the following:

Proof of Proposition 2.62 (third version). We shall prove (93) by strong induction on n:

Let $m \in \mathbb{Z}_{\geq 0}$. Assume that (93) holds for every $n \in \mathbb{Z}_{\geq 0}$ satisfying $n < m$. We must then show that (93) holds for $n = m$. In other words, we must show that $f_m \leq 2^{m-1}$ holds.

This is true if $m = 0$ (since $f_0 = 0 \leq 2^{0-1}$) and also true if $m = 1$ (since $f_1 = 1 = 2^{1-1}$ and thus $f_1 \leq 2^{1-1}$). In other words, this is true if $m \in \{0, 1\}$. Thus, for the rest of the induction step, we WLOG assume that we don't have $m \in \{0, 1\}$. Hence, $m \notin \{0, 1\}$, so that $m \in \mathbb{N} \setminus \{0, 1\} = \{2, 3, 4, \ldots\}$. Hence, $m \geq 2$.

From $m \geq 2$, we conclude that $m - 1 \geq 2 - 1 = 1 \geq 0$ and $m - 2 \geq 2 - 2 = 0$. Thus, both $m - 1$ and $m - 2$ belong to $\mathbb{N}$; therefore, f_{m-1} and f_{m-2} are well-defined.

We have $m - 1 \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ and $m - 1 < m$. Hence, (93) (applied to $n = m - 1$) yields that $f_{m-1} \leq 2^{(m-1)-1}$ (since we have assumed that (93) holds for every $n \in \mathbb{Z}_{\geq 0}$ satisfying $n < m$).

We have $m - 2 \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ and $m - 2 < m$. Hence, (93) (applied to $n = m - 2$) yields that $f_{m-2} \leq 2^{(m-2)-1}$ (since we have assumed that (93) holds for every $n \in \mathbb{Z}_{\geq 0}$ satisfying $n < m$).

We have $(m - 1) - 1 = m - 2$ and thus $2^{(m-1)-1} = 2^{m-2} = 2 \cdot 2^{(m-2)-1} \geq 2^{(m-2)-1}$ (since $2 \cdot 2^{(m-2)-1} - 2^{(m-2)-1} = 2^{(m-2)-1} \geq 0$). Hence, $2^{(m-2)-1} \leq 2^{(m-1)-1}$.

But the recursive definition of the Fibonacci sequence yields $f_m = f_{m-1} + f_{m-2}$ (since $m \geq 2$). Hence,

$$f_m = \underbrace{f_{m-1}}_{\leq 2^{(m-1)-1}} + \underbrace{f_{m-2}}_{\leq 2^{(m-2)-1} \leq 2^{(m-1)-1}} \leq 2^{(m-1)-1} + 2^{(m-1)-1} = 2 \cdot 2^{(m-1)-1} = 2^{m-1}.$$

In other words, (93) holds for $n = m$. This completes the induction step. Hence, by strong induction, we conclude that (93) holds for each $n \in \mathbb{Z}_{\geq 0}$. In other words, (93) holds for each $n \in \mathbb{N}$ (since $\mathbb{Z}_{\geq 0} = \mathbb{N}$). This completes our proof of Proposition 2.62. □

2.9. Two unexpected integralities

2.9.1. The first integrality

We shall illustrate strong induction on two further examples, which both have the form of an "unexpected integrality": A sequence of rational numbers is defined recursively by an equation that involves fractions, but it turns out that all the entries of the sequence are nevertheless integers. There is by now a whole genre of such results (see [Gale98, Chapter 1] for an introduction[67]), and many of them are connected with recent research in the theory of cluster algebras (see [Lampe13] for an introduction).

The first of these examples is the following result:

Proposition 2.64. Define a sequence $(t_0, t_1, t_2, \ldots)$ of positive rational numbers recursively by setting

$$t_0 = 1, \qquad t_1 = 1, \qquad t_2 = 1, \qquad \text{and}$$
$$t_n = \frac{1 + t_{n-1} t_{n-2}}{t_{n-3}} \qquad \text{for each } n \geq 3.$$

[67]See also [FomZel02] for a seminal research paper at a more advanced level.

(Thus,

$$t_3 = \frac{1+t_2t_1}{t_0} = \frac{1+1\cdot 1}{1} = 2;$$
$$t_4 = \frac{1+t_3t_2}{t_1} = \frac{1+2\cdot 1}{1} = 3;$$
$$t_5 = \frac{1+t_4t_3}{t_2} = \frac{1+3\cdot 2}{1} = 7;$$
$$t_6 = \frac{1+t_5t_4}{t_3} = \frac{1+7\cdot 3}{2} = 11,$$

and so on.) Then:

(a) We have $t_{n+2} = 4t_n - t_{n-2}$ for each $n \in \mathbb{Z}_{\geq 2}$.

(b) We have $t_n \in \mathbb{N}$ for each $n \in \mathbb{N}$.

Note that the sequence $(t_0, t_1, t_2, \ldots)$ in Proposition 2.64 is clearly well-defined, because the expression $\dfrac{1+t_{n-1}t_{n-2}}{t_{n-3}}$ always yields a well-defined positive rational number when $t_{n-1}, t_{n-2}, t_{n-3}$ are positive rational numbers. (In particular, the denominator t_{n-3} of this fraction is $\neq 0$ because it is positive.) In contrast, if we had set $t_n = \dfrac{1-t_{n-1}t_{n-2}}{t_{n-3}}$ instead of $t_n = \dfrac{1+t_{n-1}t_{n-2}}{t_{n-3}}$, then the sequence would **not** be well-defined (because then, we would get $t_3 = \dfrac{1-1\cdot 1}{1} = 0$ and $t_6 = \dfrac{1-t_5t_4}{t_3} = \dfrac{1-t_5t_4}{0}$, which is undefined).

Remark 2.65. The sequence $(t_0, t_1, t_2, \ldots)$ defined in Proposition 2.64 is the sequence A005246 in the OEIS (Online Encyclopedia of Integer Sequences). Its first entries are

$$t_0 = 1, \quad t_1 = 1, \quad t_2 = 1, \quad t_3 = 2, \quad t_4 = 3,$$
$$t_5 = 7, \quad t_6 = 11, \quad t_7 = 26, \quad t_8 = 41, \quad t_9 = 97.$$

Proposition 2.64 **(b)** is an instance of the *Laurent phenomenon* (see, e.g., [FomZel02, Example 3.2]).

Part **(a)** of Proposition 2.64 is proven by a (regular) induction; it is part **(b)** where strong induction comes handy:

Proof of Proposition 2.64. First, we notice that the recursive definition of the sequence $(t_0, t_1, t_2, \ldots)$ yields

$$t_3 = \frac{1+t_{3-1}t_{3-2}}{t_{3-3}} = \frac{1+t_2t_1}{t_0} = \frac{1+1\cdot 1}{1} \qquad (\text{since } t_0 = 1 \text{ and } t_1 = 1 \text{ and } t_2 = 1)$$
$$= 2.$$

Furthermore, the recursive definition of the sequence $(t_0, t_1, t_2, \ldots)$ yields

$$t_4 = \frac{1 + t_{4-1}t_{4-2}}{t_{4-3}} = \frac{1 + t_3 t_2}{t_1} = \frac{1 + 2 \cdot 1}{1} \qquad \left(\text{since } t_1 = 1 \text{ and } t_2 = 1 \text{ and } t_3 = 2\right)$$
$$= 3.$$

Thus, $t_{2+2} = t_4 = 3$. Comparing this with $4 \underbrace{t_2}_{=1} - \underbrace{t_{2-2}}_{=t_0=1} = 4 \cdot 1 - 1 = 3$, we obtain $t_{2+2} = 4t_2 - t_{2-2}$.

(a) We shall prove Proposition 2.64 **(a)** by induction on n starting at 2:

Induction base: We have already shown that $t_{2+2} = 4t_2 - t_{2-2}$. In other words, Proposition 2.64 **(a)** holds for $n = 2$. This completes the induction base.

Induction step: Let $m \in \mathbb{Z}_{\geq 2}$. Assume that Proposition 2.64 **(a)** holds for $n = m$. We must prove that Proposition 2.64 **(a)** holds for $n = m + 1$.

We have assumed that Proposition 2.64 **(a)** holds for $n = m$. In other words, we have $t_{m+2} = 4t_m - t_{m-2}$.

We have $m \in \mathbb{Z}_{\geq 2}$. Thus, m is an integer that is ≥ 2. Hence, $m \geq 2$ and thus $m + 1 \geq 2 + 1 = 3$. Thus, the recursive definition of the sequence $(t_0, t_1, t_2, \ldots)$ yields

$$t_{m+1} = \frac{1 + t_{(m+1)-1}t_{(m+1)-2}}{t_{(m+1)-3}} = \frac{1 + t_m t_{m-1}}{t_{m-2}}.$$

Multiplying this equality by t_{m-2}, we obtain $t_{m-2}t_{m+1} = 1 + t_m t_{m-1}$. In other words,

$$t_{m-2}t_{m+1} - 1 = t_m t_{m-1}. \tag{97}$$

Hence,

$$1 + \underbrace{t_{m+2}}_{=4t_m - t_{m-2}} t_{m+1} = 1 + (4t_m - t_{m-2})\, t_{m+1} = 4t_m t_{m+1} - \underbrace{(t_{m-2}t_{m+1} - 1)}_{\substack{=t_m t_{m-1} \\ \text{(by (97))}}}$$
$$= 4t_m t_{m+1} - t_m t_{m-1} = t_m (4t_{m+1} - t_{m-1}). \tag{98}$$

Also, $m + 3 \geq 3$. Thus, the recursive definition of the sequence $(t_0, t_1, t_2, \ldots)$ yields

$$t_{m+3} = \frac{1 + t_{(m+3)-1}t_{(m+3)-2}}{t_{(m+3)-3}} = \frac{1 + t_{m+2}t_{m+1}}{t_m} = \frac{1}{t_m} \underbrace{(1 + t_{m+2}t_{m+1})}_{\substack{=t_m(4t_{m+1} - t_{m-1}) \\ \text{(by (98))}}}$$
$$= \frac{1}{t_m} t_m (4t_{m+1} - t_{m-1}) = 4t_{m+1} - t_{m-1}.$$

In view of $m + 3 = (m + 1) + 2$ and $m - 1 = (m + 1) - 2$, this rewrites as $t_{(m+1)+2} = 4t_{m+1} - t_{(m+1)-2}$. In other words, Proposition 2.64 **(a)** holds for $n = m + 1$. This completes the induction step. Hence, Proposition 2.64 **(a)** is proven by induction.

(b) We shall prove Proposition 2.64 **(b)** by strong induction on n starting at 0:

Induction step: Let $m \in \mathbb{N}$. [68] Assume that Proposition 2.64 **(b)** holds for every $n \in \mathbb{N}$ satisfying $n < m$. We must now show that Proposition 2.64 **(b)** holds for $n = m$.

We have assumed that Proposition 2.64 **(b)** holds for every $n \in \mathbb{N}$ satisfying $n < m$. In other words, we have

$$t_n \in \mathbb{N} \text{ for every } n \in \mathbb{N} \text{ satisfying } n < m. \qquad (99)$$

We must now show that Proposition 2.64 **(b)** holds for $n = m$. In other words, we must show that $t_m \in \mathbb{N}$.

Recall that $(t_0, t_1, t_2, \ldots)$ is a sequence of positive rational numbers. Thus, t_m is a positive rational number.

We are in one of the following five cases:

Case 1: We have $m = 0$.

Case 2: We have $m = 1$.

Case 3: We have $m = 2$.

Case 4: We have $m = 3$.

Case 5: We have $m > 3$.

Let us first consider Case 1. In this case, we have $m = 0$. Thus, $t_m = t_0 = 1 \in \mathbb{N}$. Hence, $t_m \in \mathbb{N}$ is proven in Case 1.

Similarly, we can prove $t_m \in \mathbb{N}$ in Case 2 (using $t_1 = 1$) and in Case 3 (using $t_2 = 1$) and in Case 4 (using $t_3 = 2$). It thus remains to prove $t_m \in \mathbb{N}$ in Case 5.

So let us consider Case 5. In this case, we have $m > 3$. Thus, $m \geq 4$ (since m is an integer), so that $m - 2 \geq 4 - 2 = 2$. Thus, $m - 2$ is an integer that is ≥ 2. In other words, $m - 2 \in \mathbb{Z}_{\geq 2}$. Hence, Proposition 2.64 **(a)** (applied to $n = m - 2$) yields $t_{(m-2)+2} = 4t_{m-2} - t_{(m-2)-2}$. In view of $(m-2)+2 = m$ and $(m-2)-2 = m-4$, this rewrites as $t_m = 4t_{m-2} - t_{m-4}$.

But $m \geq 4$, so that $m - 4 \in \mathbb{N}$, and $m - 4 < m$. Hence, (99) (applied to $n = m - 4$) yields $t_{m-4} \in \mathbb{N} \subseteq \mathbb{Z}$. Similarly, $t_{m-2} \in \mathbb{Z}$.

So we know that t_{m-2} and t_{m-4} are both integers (since $t_{m-2} \in \mathbb{Z}$ and $t_{m-4} \in \mathbb{Z}$). Hence, $4t_{m-2} - t_{m-4}$ is an integer as well. In other words, t_m is an integer (because $t_m = 4t_{m-2} - t_{m-4}$). Since t_m is positive, we thus conclude that t_m is a positive integer. Hence, $t_m \in \mathbb{N}$. This shows that $t_m \in \mathbb{N}$ in Case 5.

We now have proven $t_m \in \mathbb{N}$ in each of the five Cases 1, 2, 3, 4 and 5. Since these five Cases cover all possibilities, we thus conclude that $t_m \in \mathbb{N}$ always holds. In other words, Proposition 2.64 **(b)** holds for $n = m$. This completes the induction step. Thus, Proposition 2.64 **(b)** is proven by strong induction. □

2.9.2. The second integrality

Our next example of an "unexpected integrality" is the following fact:

[68] In order to match the notations used in Theorem 2.60, we should be saying "Let $m \in \mathbb{Z}_{\geq 0}$" here, rather than "Let $m \in \mathbb{N}$". But of course, this amounts to the same thing, since $\mathbb{N} = \mathbb{Z}_{\geq 0}$.

Proposition 2.66. Fix a positive integer r. Define a sequence $(b_0, b_1, b_2, \ldots)$ of positive rational numbers recursively by setting

$$b_0 = 1, \qquad b_1 = 1, \qquad \text{and}$$
$$b_n = \frac{b_{n-1}^r + 1}{b_{n-2}} \qquad \text{for each } n \geq 2.$$

(Thus,

$$b_2 = \frac{b_1^r + 1}{b_0} = \frac{1^r + 1}{1} = 2;$$
$$b_3 = \frac{b_2^r + 1}{b_1} = \frac{2^r + 1}{1} = 2^r + 1;$$
$$b_4 = \frac{b_3^r + 1}{b_2} = \frac{(2^r + 1)^r + 1}{2},$$

and so on.) Then:

(a) We have $b_n \in \mathbb{N}$ for each $n \in \mathbb{N}$.

(b) If $r \geq 2$, then $b_n \mid b_{n-2} + b_{n+2}$ for each $n \in \mathbb{Z}_{\geq 2}$.

Remark 2.67. If $r = 1$, then the sequence $(b_0, b_1, b_2, \ldots)$ defined in Proposition 2.66 is

$$(1, 1, 2, 3, 2, 1, 1, 2, 3, 2, 1, 1, 2, 3, 2, \ldots)$$

(this is a periodic sequence, which consists of the five terms $1, 1, 2, 3, 2$ repeated over and over); this can easily be proven by strong induction. Despite its simplicity, this sequence is the sequence A076839 in the OEIS.

If $r = 2$, then the sequence $(b_0, b_1, b_2, \ldots)$ defined in Proposition 2.66 is

$$(1, f_1, f_3, f_5, f_7, \ldots) = (1, 1, 2, 5, 13, 34, 89, 233, 610, 1597, \ldots)$$

consisting of all Fibonacci numbers at odd positions (i.e., Fibonacci numbers of the form f_{2n-1} for $n \in \mathbb{N}$) with an extra 1 at the front. This, again, can be proven by induction. Also, this sequence satisfies the recurrence relation $b_n = 3b_{n-1} - b_{n-2}$ for all $n \geq 2$. This is the sequence A001519 in the OEIS.

If $r = 3$, then the sequence $(b_0, b_1, b_2, \ldots)$ defined in Proposition 2.66 is

$$(1, 1, 2, 9, 365, 5403014, 432130991537958813, \ldots);$$

its entries grow so fast that the next entry would need a separate line. This is the sequence A003818 in the OEIS. Unlike the cases of $r = 1$ and $r = 2$, not much can be said about this sequence, other than what has been said in Proposition 2.66.

Proposition 2.66 **(a)** is an instance of the *Laurent phenomenon for cluster algebras* (see, e.g., [FomZel01, Example 2.5]; also, see [Marsh13] and [FoWiZe16] for expositions). See also [MusPro07] for a study of the specific recurrence equation from Proposition 2.66 (actually, a slightly more general equation).

Before we prove Proposition 2.66, let us state an auxiliary fact:

Lemma 2.68. Let $r \in \mathbb{N}$. For every nonzero $x \in \mathbb{Q}$, we set $H(x) = \dfrac{(x+1)^r - 1}{x}$. Then, $H(x) \in \mathbb{Z}$ whenever x is a nonzero integer.

Proof of Lemma 2.68. Let x be a nonzero integer. Then, $x \mid (x+1) - 1$ (because $(x+1) - 1 = x$). In other words, $x + 1 \equiv 1 \mod x$ (by the definition of "congruent"). Hence, Proposition 2.22 (applied to $a = x+1$, $b = 1$, $n = x$ and $k = r$) shows that $(x+1)^r \equiv 1^r = 1 \mod x$. In other words, $(x+1)^r - 1$ is divisible by x. In other words, $\dfrac{(x+1)^r - 1}{x}$ is an integer. In other words, $\dfrac{(x+1)^r - 1}{x} \in \mathbb{Z}$. Thus, $H(x) = \dfrac{(x+1)^r - 1}{x} \in \mathbb{Z}$. This proves Lemma 2.68. □

Proof of Proposition 2.66. First, we notice that the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields

$$\begin{aligned} b_2 &= \frac{b_{2-1}^r + 1}{b_{2-2}} = \frac{b_1^r + 1}{b_0} = \frac{1^r + 1}{1} \qquad (\text{since } b_0 = 1 \text{ and } b_1 = 1) \\ &= \frac{1+1}{1} \qquad (\text{since } 1^r = 1) \\ &= 2. \end{aligned}$$

Furthermore, the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields

$$\begin{aligned} b_3 &= \frac{b_{3-1}^r + 1}{b_{3-2}} = \frac{b_2^r + 1}{b_1} = \frac{2^r + 1}{1} \qquad (\text{since } b_1 = 1 \text{ and } b_2 = 2) \\ &= 2^r + 1. \end{aligned}$$

For every nonzero $x \in \mathbb{Q}$, we set $H(x) = \dfrac{(x+1)^r - 1}{x}$.

For every integer $m \geq 1$, we have

$$b_m^r + 1 = b_{m+1} b_{m-1}. \tag{100}$$

[*Proof of (100):* Let $m \geq 1$ be an integer. From $m \geq 1$, we obtain $m + 1 \geq 1 + 1 = 2$. Hence, the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields

$$b_{m+1} = \frac{b_{(m+1)-1}^r + 1}{b_{(m+1)-2}} = \frac{b_m^r + 1}{b_{m-1}}.$$

Multiplying both sides of this equality by b_{m-1}, we obtain $b_{m+1} b_{m-1} = b_m^r + 1$. This proves (100).]

Let us first prove the following observation:

Observation 1: Each integer $n \geq 2$ satisfies $b_{n+2} = b_{n-2} b_{n+1}^r - b_n^{r-1} H(b_n^r)$.

[*Proof of Observation 1:* Let $n \geq 2$ be an integer. Thus, $n \geq 2 \geq 1$. Thus, (100) (applied to $m = n$) yields

$$b_n^r + 1 = b_{n+1} b_{n-1}. \tag{101}$$

On the other hand, $n + 1 \geq n \geq 2 \geq 1$. Hence, (100) (applied to $m = n + 1$) yields

$$b_{n+1}^r + 1 = b_{(n+1)+1} b_{(n+1)-1} = b_{n+2} b_n.$$

Hence,

$$b_{n+1}^r = b_{n+2} b_n - 1. \tag{102}$$

Also, $n - 1 \geq 1$ (since $n \geq 2 = 1 + 1$). Hence, (100) (applied to $m = n - 1$) yields

$$b_{n-1}^r + 1 = b_{(n-1)+1} b_{(n-1)-1} = b_n b_{n-2}.$$

Hence,

$$b_{n-1}^r = b_n b_{n-2} - 1. \tag{103}$$

But b_n is a positive rational number (since $(b_0, b_1, b_2, \ldots)$ is a sequence of positive rational numbers). Thus, b_n^r is also a positive rational number. Hence, $b_n^r \in \mathbb{Q}$ is nonzero. The definition of $H(b_n^r)$ yields $H(b_n^r) = \dfrac{(b_n^r + 1)^r - 1}{b_n^r}$; therefore,

$$\begin{aligned}
b_n^{r-1} \underbrace{H(b_n^r)}_{= \frac{(b_n^r+1)^r - 1}{b_n^r}} &= b_n^{r-1} \cdot \frac{(b_n^r + 1)^r - 1}{b_n^r} = \underbrace{\frac{b_n^{r-1}}{b_n^r}}_{= \frac{1}{b_n}} \cdot \left(\Bigg(\underbrace{b_n^r + 1}_{\substack{= b_{n+1} b_{n-1} \\ \text{(by (101))}}} \Bigg)^r - 1 \right) \\
&= \frac{1}{b_n} \cdot \left(\underbrace{(b_{n+1} b_{n-1})^r}_{= b_{n+1}^r b_{n-1}^r} - 1 \right) = \frac{1}{b_n} \cdot \left(\underbrace{b_{n+1}^r}_{\substack{= b_{n+2} b_n - 1 \\ \text{(by (102))}}} \underbrace{b_{n-1}^r}_{\substack{= b_n b_{n-2} - 1 \\ \text{(by (103))}}} - 1 \right) \\
&= \frac{1}{b_n} \cdot \underbrace{((b_{n+2} b_n - 1)(b_n b_{n-2} - 1) - 1)}_{= b_n (b_n b_{n+2} b_{n-2} - b_{n+2} - b_{n-2})} \\
&= \frac{1}{b_n} \cdot b_n (b_n b_{n+2} b_{n-2} - b_{n+2} - b_{n-2}) \\
&= b_n b_{n+2} b_{n-2} - b_{n+2} - b_{n-2} = b_{n-2} \underbrace{(b_{n+2} b_n - 1)}_{\substack{= b_{n+1}^r \\ \text{(by (102))}}} - b_{n+2} \\
&= b_{n-2} b_{n+1}^r - b_{n+2}.
\end{aligned}$$

Solving this equation for b_{n+2}, we obtain $b_{n+2} = b_{n-2} b_{n+1}^r - b_n^{r-1} H(b_n^r)$. This proves Observation 1.]

(a) We shall prove Proposition 2.66 **(a)** by strong induction on n:

Induction step: Let $m \in \mathbb{N}$. Assume that Proposition 2.66 **(a)** holds for every $n \in \mathbb{N}$ satisfying $n < m$. We must now prove that Proposition 2.66 **(a)** holds for $n = m$.

We have assumed that Proposition 2.66 **(a)** holds for every $n \in \mathbb{N}$ satisfying $n < m$. In other words, we have

$$b_n \in \mathbb{N} \text{ for every } n \in \mathbb{N} \text{ satisfying } n < m. \tag{104}$$

We must now show that Proposition 2.66 **(a)** holds for $n = m$. In other words, we must show that $b_m \in \mathbb{N}$.

Recall that $(b_0, b_1, b_2, \ldots)$ is a sequence of positive rational numbers. Thus, b_m is a positive rational number.

We are in one of the following five cases:

Case 1: We have $m = 0$.

Case 2: We have $m = 1$.

Case 3: We have $m = 2$.

Case 4: We have $m = 3$.

Case 5: We have $m > 3$.

Let us first consider Case 1. In this case, we have $m = 0$. Thus, $b_m = b_0 = 1 \in \mathbb{N}$. Hence, $b_m \in \mathbb{N}$ is proven in Case 1.

Similarly, we can prove $b_m \in \mathbb{N}$ in Case 2 (using $b_1 = 1$) and in Case 3 (using $b_2 = 2$) and in Case 4 (using $b_3 = 2^r + 1$). It thus remains to prove $b_m \in \mathbb{N}$ in Case 5.

So let us consider Case 5. In this case, we have $m > 3$. Thus, $m \geq 4$ (since m is an integer), so that $m - 2 \geq 4 - 2 = 2$. Hence, Observation 1 (applied to $n = m - 2$) yields $b_{(m-2)+2} = b_{(m-2)-2} b^r_{(m-2)+1} - b^{r-1}_{m-2} H\left(b^r_{m-2}\right)$. In view of $(m-2)+2 = m$ and $(m-2)-2 = m-4$ and $(m-2)+1 = m-1$, this rewrites as

$$b_m = b_{m-4} b^r_{m-1} - b^{r-1}_{m-2} H\left(b^r_{m-2}\right). \tag{105}$$

But $m - 2 \in \mathbb{N}$ (since $m \geq 4 \geq 2$) and $m - 2 < m$. Hence, (104) (applied to $n = m - 2$) yields $b_{m-2} \in \mathbb{N} \subseteq \mathbb{Z}$. Also, b_{m-2} is a positive rational number (since $(b_0, b_1, b_2, \ldots)$ is a sequence of positive rational numbers) and thus a positive integer (since $b_{m-2} \in \mathbb{N}$), hence a nonzero integer. Thus, b^r_{m-2} is a nonzero integer as well. Therefore, Lemma 2.68 (applied to $x = b^r_{m-2}$) shows that $H\left(b^r_{m-2}\right) \in \mathbb{Z}$. In other words, $H\left(b^r_{m-2}\right)$ is an integer. Also, $r - 1 \geq 0$ (since $r \geq 1$), and thus $r - 1 \in \mathbb{N}$. Hence, b^{r-1}_{m-2} is an integer (since b_{m-2} is an integer).

Also, $m - 4 \in \mathbb{N}$ (since $m \geq 4$) and $m - 4 < m$. Hence, (104) (applied to $n = m - 4$) yields $b_{m-4} \in \mathbb{N} \subseteq \mathbb{Z}$. In other words, b_{m-4} is an integer.

Similarly, b_{m-1} is an integer. Thus, b^r_{m-1} is an integer.

We now know that the four numbers b_{m-4}, b^r_{m-1}, b^{r-1}_{m-2} and $H\left(b^r_{m-2}\right)$ are integers. Thus, the number $b_{m-4} b^r_{m-1} - b^{r-1}_{m-2} H\left(b^r_{m-2}\right)$ also is an integer (since it is obtained from these four numbers by multiplication and subtraction). In view of (105), this rewrites as follows: The number b_m is an integer. Since b_m is positive, we thus conclude that b_m is a positive integer. Hence, $b_m \in \mathbb{N}$. This shows that $b_m \in \mathbb{N}$ in Case 5.

We now have proven $b_m \in \mathbb{N}$ in each of the five Cases 1, 2, 3, 4 and 5. Thus, $b_m \in \mathbb{N}$ always holds. In other words, Proposition 2.66 **(a)** holds for $n = m$. This completes the induction step. Thus, Proposition 2.66 **(a)** is proven by strong induction.

(b) Assume that $r \geq 2$. We must prove that $b_n \mid b_{n-2} + b_{n+2}$ for each $n \in \mathbb{Z}_{\geq 2}$.

So let $n \in \mathbb{Z}_{\geq 2}$. We must show that $b_n \mid b_{n-2} + b_{n+2}$.

From $n \in \mathbb{Z}_{\geq 2}$, we obtain $n \geq 2$, so that $n - 2 \in \mathbb{N}$.

Proposition 2.66 **(a)** (applied to $n-2$ instead of n) yields $b_{n-2} \in \mathbb{N}$. Similarly, $b_n \in \mathbb{N}$ and $b_{n+1} \in \mathbb{N}$ and $b_{n+2} \in \mathbb{N}$. Thus, all of b_{n-2}, b_{n+1}, b_n are b_{n+2} are integers.

But b_n is a positive rational number (since $(b_0, b_1, b_2, \ldots)$ is a sequence of positive rational numbers), and therefore a positive integer (since b_n is an integer). Hence, b_n^r is a positive integer, and thus a nonzero integer. Therefore, Lemma 2.68 (applied to $x = b_n^r$) shows that $H(b_n^r) \in \mathbb{Z}$. In other words, $H(b_n^r)$ is an integer.

We have $n + 2 \geq 2$. Hence, the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields $b_{n+2} = \dfrac{b_{(n+2)-1}^r + 1}{b_{(n+2)-2}} = \dfrac{b_{n+1}^r + 1}{b_n}$. Multiplying this equality by b_n, we obtain

$$b_n b_{n+2} = b_{n+1}^r + 1. \tag{106}$$

We have $r - 2 \in \mathbb{N}$ (since $r \geq 2$) and $b_n \in \mathbb{N}$. Hence, b_n^{r-2} is an integer.

Observation 1 yields $b_{n+2} = b_{n-2} b_{n+1}^r - b_n^{r-1} H(b_n^r)$. Thus,

$$\begin{aligned}
&\underbrace{b_{n+2}}_{= b_{n-2} b_{n+1}^r - b_n^{r-1} H(b_n^r)} + b_{n-2} \\
&= b_{n-2} b_{n+1}^r - b_n^{r-1} H(b_n^r) + b_{n-2} = \underbrace{b_{n-2} b_{n+1}^r + b_{n-2}}_{= b_{n-2}(b_{n+1}^r + 1)} - b_n^{r-1} H(b_n^r) \\
&= b_{n-2} \underbrace{(b_{n+1}^r + 1)}_{\substack{= b_n b_{n+2} \\ \text{(by (106))}}} - \underbrace{b_n^{r-1}}_{= b_n b_n^{r-2}} H(b_n^r) = b_{n-2} b_n b_{n+2} - b_n b_n^{r-2} H(b_n^r) \\
&= b_n \left(b_{n-2} b_{n+2} - b_n^{r-2} H(b_n^r) \right).
\end{aligned} \tag{107}$$

But $b_{n-2} b_{n+2} - b_n^{r-2} H(b_n^r)$ is an integer (because b_{n-2}, b_{n+2}, b_n^{r-2} and $H(b_n^r)$ are integers). Denote this integer by z. Thus, $z = b_{n-2} b_{n+2} - b_n^{r-2} H(b_n^r)$. Since b_n and

z are integers, we have

$$\begin{aligned} b_n \mid b_n \underbrace{z}_{=b_{n-2}b_{n+2}-b_n^{r-2}H(b_n^r)} &= b_n \left(b_{n-2}b_{n+2} - b_n^{r-2} H\left(b_n^r\right)\right) = b_{n+2} + b_{n-2} \qquad \text{(by (107))} \\ &= b_{n-2} + b_{n+2}. \end{aligned}$$

This proves Proposition 2.66 **(b)**. □

For a (slightly) different proof of Proposition 2.66, see `http://artofproblemsolving.com/community/c6h428645p3705719` .

Remark 2.69. You might wonder what happens if we replace "$b_{n-1}^r + 1$" by "$b_{n-1}^r + q$" in Proposition 2.66, where q is some fixed nonnegative integer. The answer turns out to be somewhat disappointing in general: For example, if we set $r = 3$ and $q = 2$, then our sequence $(b_0, b_1, b_2, \ldots)$ begins with

$$b_0 = 1, \qquad b_1 = 1, \qquad b_2 = \frac{1^3 + 2}{1} = 3,$$
$$b_3 = \frac{3^3 + 2}{1} = 29, \qquad b_4 = \frac{29^3 + 2}{3} = \frac{24\,391}{3},$$

at which point it becomes clear that $b_n \in \mathbb{N}$ no longer holds for all $n \in \mathbb{N}$. The same happens for $r = 4, 5, \ldots, 11$ (though the first n that violates $b_n \in \mathbb{N}$ is not always 4); this makes me suspect that it also happens for all $r > 2$.

However, $b_n \in \mathbb{N}$ still holds for all $n \in \mathbb{N}$ when $r = 2$. This follows from Exercise 2.1 below.

Exercise 2.1. Fix a nonnegative integer q. Define a sequence $(b_0, b_1, b_2, \ldots)$ of positive rational numbers recursively by setting

$$b_0 = 1, \qquad b_1 = 1, \qquad \text{and}$$
$$b_n = \frac{b_{n-1}^2 + q}{b_{n-2}} \qquad \text{for each } n \geq 2.$$

(Thus,

$$\begin{aligned} b_2 &= \frac{b_1^2 + q}{b_0} = \frac{1^2 + q}{1} = q + 1; \\ b_3 &= \frac{b_2^2 + q}{b_1} = \frac{(q+1)^2 + q}{1} = q^2 + 3q + 1; \\ b_4 &= \frac{b_3^2 + q}{b_2} = \frac{\left(q^2 + 3q + 1\right)^2 + q}{q + 1} = q^3 + 5q^2 + 6q + 1; \end{aligned}$$

and so on.) Prove that:

(a) We have $b_n = (q + 2) b_{n-1} - b_{n-2}$ for each $n \in \mathbb{Z}_{\geq 2}$.

(b) We have $b_n \in \mathbb{N}$ for each $n \in \mathbb{N}$.

2.10. Strong induction on a derived quantity: Bezout's theorem

2.10.1. Strong induction on a derived quantity

In Section 2.5, we have seen how to use induction on a variable that does not explicitly appear in the claim. In the current section, we shall show the same for strong induction. This time, the fact that we shall be proving is the following:

Theorem 2.70. Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$. Then, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax + by$ and $g \mid a$ and $g \mid b$.

Example 2.71. (a) If $a = 3$ and $b = 5$, then Theorem 2.70 says that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = 3x + 5y$ and $g \mid 3$ and $g \mid 5$. And indeed, it is easy to find such g, x and y: for example, $g = 1$, $x = -3$ and $y = 2$ will do (since $1 = 3(-3) + 5 \cdot 2$ and $1 \mid 3$ and $1 \mid 5$).

(b) If $a = 4$ and $b = 6$, then Theorem 2.70 says that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = 4x + 6y$ and $g \mid 4$ and $g \mid 6$. And indeed, it is easy to find such g, x and y: for example, $g = 2$, $x = -1$ and $y = 1$ will do (since $2 = 4(-1) + 6 \cdot 1$ and $2 \mid 4$ and $2 \mid 6$).

Theorem 2.70 is one form of *Bezout's theorem for integers*, and its real significance might not be clear at this point; it becomes important when the greatest common divisor of two integers is studied. For now, we observe that the g in Theorem 2.70 is obviously a common divisor of a and b (that is, an integer that divides both a and b); but it is also divisible by every common divisor of a and b (because of Proposition 2.7).

Let us now focus on the proof of Theorem 2.70. It is natural to try proving it by induction (or perhaps strong induction) on a or on b, but neither option leads to success. It may feel like "induction on a and on b at the same time" could help, and this is indeed a viable approach[69]. But there is a simpler and shorter method available: strong induction on $a + b$. As in Section 2.5, the way to formalize such a

[69] Of course, it needs to be done right: An induction proof always requires choosing **one** variable to do induction on; but it is possible to nest an induction proof inside the induction step (or inside the induction base) of a different induction proof. For example, imagine that we are trying to prove that

$$ab = ba \qquad \text{for any } a \in \mathbb{N} \text{ and } b \in \mathbb{N}.$$

We can prove this by induction on a. More precisely, for each $a \in \mathbb{N}$, we let $\mathcal{A}(a)$ be the statement ($ab = ba$ for all $b \in \mathbb{N}$). We then prove $\mathcal{A}(a)$ by induction on a. In the induction step, we fix $m \in \mathbb{N}$, and we assume that $\mathcal{A}(m)$ holds; we now need to prove that $\mathcal{A}(m+1)$ holds. In other words, we need to prove that $(m+1)b = b(m+1)$ for all $b \in \mathbb{N}$. We can now prove this statement by induction on b (although there are easier options, of course). Thus, the induction proof of this statement happens inside the induction step of another induction proof. This nesting of induction proofs is legitimate (and even has a name: it is called *double induction*), but tends to be rather confusing (just think about what the sentence "The induction base is complete" means: is it about the induction base of the first induction proof, or that of the second?), and is best avoided when possible.

strong induction is by introducing auxiliary statements $\mathcal{A}(n)$, which say as much as "Theorem 2.70 holds under the requirement that $a+b=n$":

Proof of Theorem 2.70. First of all, let us forget that we fixed a and b. So we want to prove that if $a \in \mathbb{N}$ and $b \in \mathbb{N}$, then there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax + by$ and $g \mid a$ and $g \mid b$.

For each $n \in \mathbb{N}$, we let $\mathcal{A}(n)$ be the statement

$$\left(\begin{array}{c} \text{if } a \in \mathbb{N} \text{ and } b \in \mathbb{N} \text{ satisfy } a+b=n, \text{ then there exist } g \in \mathbb{N},\ x \in \mathbb{Z} \\ \text{and } y \in \mathbb{Z} \text{ such that } g = ax+by \text{ and } g \mid a \text{ and } g \mid b \end{array}\right). \qquad (108)$$

We claim that $\mathcal{A}(n)$ holds for all $n \in \mathbb{N}$.

Indeed, let us prove this by strong induction on n starting at 0:

Induction step: Let $m \in \mathbb{N}$. Assume that

$$(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{N} \text{ satisfying } n < m). \qquad (109)$$

We must then show that $\mathcal{A}(m)$ holds.

To do this, we shall prove the following claim:

> *Claim 1:* Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$ satisfy $a+b=m$. Then, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax+by$ and $g \mid a$ and $g \mid b$.

Before we prove Claim 1, let us show a slightly weaker version of it, in which we rename a and b as u and v and add the assumption that $u \geq v$:

> *Claim 2:* Let $u \in \mathbb{N}$ and $v \in \mathbb{N}$ satisfy $u+v=m$ and $u \geq v$. Then, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ux+vy$ and $g \mid u$ and $g \mid v$.

[*Proof of Claim 2:* We are in one of the following two cases:

Case 1: We have $v = 0$.

Case 2: We have $v \neq 0$.

Let us first consider Case 1. In this case, we have $v = 0$. Hence, $v = 0 = 0u$, so that $u \mid v$. Also, $u \cdot 1 + v \cdot 0 = u$. Thus, $u = u \cdot 1 + v \cdot 0$ and $u \mid u$ and $u \mid v$. Hence, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ux+vy$ and $g \mid u$ and $g \mid v$ (namely, $g = u$, $x = 1$ and $y = 0$). Thus, Claim 2 is proven in Case 1.

Let us now consider Case 2. In this case, we have $v \neq 0$. Hence, $v > 0$ (since $v \in \mathbb{N}$). Thus, $u + v > u + 0 = u$, so that $u < u + v = m$. Hence, (109) (applied to $n = u$) yields that $\mathcal{A}(u)$ holds. In other words,

$$\left(\begin{array}{c} \text{if } a \in \mathbb{N} \text{ and } b \in \mathbb{N} \text{ satisfy } a+b=u, \text{ then there exist } g \in \mathbb{N},\ x \in \mathbb{Z} \\ \text{and } y \in \mathbb{Z} \text{ such that } g = ax+by \text{ and } g \mid a \text{ and } g \mid b \end{array}\right) \qquad (110)$$

(because this is what the statement $\mathcal{A}(u)$ says).

Also, $u - v \in \mathbb{N}$ (since $u \geq v$) and $(u-v)+v = u$. Hence, (110) (applied to $a = u - v$ and $b = v$) shows that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that

$g = (u - v)x + vy$ and $g \mid u - v$ and $g \mid v$. Consider these g, x and y, and denote them by g', x' and y'. Thus, g' is an element of $\mathbb{N}$, and x' and y' are elements of $\mathbb{Z}$ satisfying $g' = (u - v)x' + vy'$ and $g' \mid u - v$ and $g' \mid v$.

Now, we have $g' \mid u - v$; in other words, $u \equiv v \mod g'$. Also, $g' \mid v$; in other words, $v \equiv 0 \mod g'$. Hence, $u \equiv v \equiv 0 \mod g'$, so that $u \equiv 0 \mod g'$. In other words, $g' \mid u$. Furthermore,

$$g' = (u - v)x' + vy' = ux' - vx' + vy' = ux' + v(y' - x').$$

Hence, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ux + vy$ and $g \mid u$ and $g \mid v$ (namely, $g = g'$, $x = x'$ and $y = y' - x'$). Thus, Claim 2 is proven in Case 2.

We have now proven Claim 2 in each of the two Cases 1 and 2. Thus, Claim 2 always holds (since Cases 1 and 2 cover all possibilities).]

Now, we can prove Claim 1 as well:

[*Proof of Claim 1:* We are in one of the following two cases:

Case 1: We have $a \geq b$.

Case 2: We have $a < b$.

Let us first consider Case 1. In this case, we have $a \geq b$. Hence, Claim 2 (applied to $u = a$ and $v = b$) shows that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax + by$ and $g \mid a$ and $g \mid b$. Thus, Claim 1 is proven in Case 1.

Let us next consider Case 2. In this case, we have $a < b$. Hence, $a \leq b$, so that $b \geq a$. Also, $b + a = a + b = m$. Hence, Claim 2 (applied to $u = b$ and $v = a$) shows that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = bx + ay$ and $g \mid b$ and $g \mid a$. Consider these g, x and y, and denote them by g', x' and y'. Thus, g' is an element of $\mathbb{N}$, and x' and y' are elements of $\mathbb{Z}$ satisfying $g' = bx' + ay'$ and $g' \mid b$ and $g' \mid a$. Now, $g' = bx' + ay' = ay' + bx'$. Hence, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax + by$ and $g \mid a$ and $g \mid b$ (namely, $g = g'$, $x = y'$ and $y = x'$). Thus, Claim 1 is proven in Case 2.

We have now proven Claim 1 in each of the two Cases 1 and 2. Thus, Claim 1 always holds (since Cases 1 and 2 cover all possibilities).]

But $\mathcal{A}(m)$ is defined as the statement

$$\left(\begin{array}{c}\text{if } a \in \mathbb{N} \text{ and } b \in \mathbb{N} \text{ satisfy } a + b = m \text{, then there exist } g \in \mathbb{N},\ x \in \mathbb{Z} \\ \text{and } y \in \mathbb{Z} \text{ such that } g = ax + by \text{ and } g \mid a \text{ and } g \mid b\end{array}\right).$$

Thus, $\mathcal{A}(m)$ is precisely Claim 1. Hence, $\mathcal{A}(m)$ holds (since Claim 1 holds). This completes the induction step. Thus, we have proven by strong induction that $\mathcal{A}(n)$ holds for all $n \in \mathbb{N}$. In other words, the statement (108) holds for all $n \in \mathbb{N}$ (since this statement is precisely $\mathcal{A}(n)$).

Now, let $a \in \mathbb{N}$ and $b \in \mathbb{N}$. Then, $a + b \in \mathbb{N}$. Hence, we can apply (108) to $n = a + b$ (since $a + b = a + b$). We thus conclude that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax + by$ and $g \mid a$ and $g \mid b$. This proves Theorem 2.70. $\square$

2.10.2. Conventions for writing proofs by strong induction on derived quantities

Let us take a closer look at the proof we just gave. The statement $\mathcal{A}(n)$ that we defined was unsurprising: It simply says that Theorem 2.70 holds under the condition that $a+b=n$. Thus, by introducing $\mathcal{A}(n)$, we have "sliced" Theorem 2.70 into a sequence of statements $\mathcal{A}(0), \mathcal{A}(1), \mathcal{A}(2), \ldots$, which then allowed us to prove these statements by strong induction on n even though no "n" appeared in Theorem 2.70 itself. This strong induction can be simply called a "strong induction on $a+b$". More generally:

Convention 2.72. Let $\mathcal{B}$ be a logical statement that involves some variables $v_1, v_2, v_3, \ldots$. (For example, $\mathcal{B}$ can be the statement of Theorem 2.70; then, these variables are a and b.)

Let $g \in \mathbb{Z}$. (This g has nothing to do with the g from Theorem 2.70.)

Let q be some expression (involving the variables $v_1, v_2, v_3, \ldots$ or some of them) that has the property that whenever the variables $v_1, v_2, v_3, \ldots$ satisfy the assumptions of $\mathcal{B}$, the expression q evaluates to some element of $\mathbb{Z}_{\geq g}$. (For example, if $\mathcal{B}$ is the statement of Theorem 2.70 and $g=0$, then q can be the expression $a+b$, because $a+b \in \mathbb{N} = \mathbb{Z}_{\geq 0}$ whenever a and b are as in Theorem 2.70.)

Assume that you want to prove the statement $\mathcal{B}$. Then, you can proceed as follows: For each $n \in \mathbb{Z}_{\geq g}$, define $\mathcal{A}(n)$ to be the statement saying that[70]

$$(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = n).$$

Then, prove $\mathcal{A}(n)$ by strong induction on n starting at g. Thus:

- The *induction step* consists in fixing $m \in \mathbb{Z}_{\geq g}$, and showing that if

  $$(\mathcal{A}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < m), \tag{111}$$

  then

  $$(\mathcal{A}(m) \text{ holds}). \tag{112}$$

  In other words, it consists in fixing $m \in \mathbb{Z}_{\geq g}$, and showing that if

  $$(\text{the statement } \mathcal{B} \text{ holds under the condition that } q < m), \tag{113}$$

  then

  $$(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = m). \tag{114}$$

  (Indeed, the previous two sentences are equivalent, because of the logical equivalences

$$\begin{aligned}
&\left(\mathcal{A}\left(n\right) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < m\right) \\
\Longleftrightarrow\ &\left(\begin{array}{c}\left(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = n\right) \\ \text{holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < m\end{array}\right) \\
&\qquad\left(\begin{array}{c}\text{since the statement } \mathcal{A}\left(n\right) \text{ is defined as} \\ \left(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = n\right)\end{array}\right) \\
\Longleftrightarrow\ &\left(\text{the statement } \mathcal{B} \text{ holds under the condition that } q < m\right)
\end{aligned}$$

and

$$\begin{aligned}
&\left(\mathcal{A}\left(m\right) \text{ holds}\right) \\
\Longleftrightarrow\ &\left(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = m\right) \\
&\qquad\left(\begin{array}{c}\text{since the statement } \mathcal{A}\left(m\right) \text{ is defined as} \\ \left(\text{the statement } \mathcal{B} \text{ holds under the condition that } q = m\right)\end{array}\right).
\end{aligned}$$

)

In practice, this induction step will usually be organized as follows: We fix $m \in \mathbb{Z}_{\geq g}$, then we assume that the statement $\mathcal{B}$ holds under the condition that $q < m$ (this is the induction hypothesis), and then we prove that the statement $\mathcal{B}$ holds under the condition that $q = m$.

Once this induction proof is finished, it immediately follows that the statement $\mathcal{B}$ always holds (because the induction proof has shown that, whatever $n \in \mathbb{Z}_{\geq g}$ is, the statement $\mathcal{B}$ holds under the condition that $q = n$).

This strategy of proof is called "strong induction on q" (or "strong induction over q"). Once you have specified what q is, you don't need to explicitly define $\mathcal{A}\left(n\right)$, nor do you ever need to mention n.

Using this convention, we can rewrite our above proof of Theorem 2.70 as follows (remembering once again that $\mathbb{Z}_{\geq 0} = \mathbb{N}$):

Proof of Theorem 2.70 (second version). Let us prove Theorem 2.70 by strong induction on $a + b$ starting at 0:

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.70 holds under the condition that $a + b < m$. We must then show that Theorem 2.70 holds under the condition that $a + b = m$. This is tantamount to proving the following claim:

> *Claim 1:* Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$ satisfy $a + b = m$. Then, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ax + by$ and $g \mid a$ and $g \mid b$.

[70] We assume that no variable named "n" appears in the statement $\mathcal{B}$; otherwise, we need a different letter for our new variable in order to avoid confusion.

Before we prove Claim 1, let us show a slightly weaker version of it, in which we rename a and b as u and v and add the assumption that $u \geq v$:

> *Claim 2:* Let $u \in \mathbb{N}$ and $v \in \mathbb{N}$ satisfy $u + v = m$ and $u \geq v$. Then, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ux + vy$ and $g \mid u$ and $g \mid v$.

[*Proof of Claim 2:* We are in one of the following two cases:

Case 1: We have $v = 0$.

Case 2: We have $v \neq 0$.

Let us first consider Case 1. In this case, we have $v = 0$. Hence, $v = 0 = 0u$, so that $u \mid v$. Also, $u \cdot 1 + v \cdot 0 = u$. Thus, $u = u \cdot 1 + v \cdot 0$ and $u \mid u$ and $u \mid v$. Hence, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ux + vy$ and $g \mid u$ and $g \mid v$ (namely, $g = u$, $x = 1$ and $y = 0$). Thus, Claim 2 is proven in Case 1.

Let us now consider Case 2. In this case, we have $v \neq 0$. Hence, $v > 0$ (since $v \in \mathbb{N}$). Thus, $u + v > u + 0 = u$, so that $u < u + v = m$. Also, $u - v \in \mathbb{N}$ (since $u \geq v$) and $(u - v) + v = u$.

But we assumed that Theorem 2.70 holds under the condition that $a + b < m$. Thus, we can apply Theorem 2.70 to $a = u - v$ and $b = v$ (since $u - v \in \mathbb{N}$ and $(u - v) + v = u < m$). We thus conclude that there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = (u - v)\,x + vy$ and $g \mid u - v$ and $g \mid v$. Consider these g, x and y, and denote them by g', x' and y'. Thus, g' is an element of $\mathbb{N}$, and x' and y' are elements of $\mathbb{Z}$ satisfying $g' = (u - v)\,x' + vy'$ and $g' \mid u - v$ and $g' \mid v$.

Now, we have $g' \mid u - v$; in other words, $u \equiv v \mod g'$. Also, $g' \mid v$; in other words, $v \equiv 0 \mod g'$. Hence, $u \equiv v \equiv 0 \mod g'$, so that $u \equiv 0 \mod g'$. In other words, $g' \mid u$. Furthermore,

$$g' = (u - v)\,x' + vy' = ux' - vx' + vy' = ux' + v\left(y' - x'\right).$$

Hence, there exist $g \in \mathbb{N}$, $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $g = ux + vy$ and $g \mid u$ and $g \mid v$ (namely, $g = g'$, $x = x'$ and $y = y' - x'$). Thus, Claim 2 is proven in Case 2.

We have now proven Claim 2 in each of the two Cases 1 and 2. Thus, Claim 2 always holds (since Cases 1 and 2 cover all possibilities).]

Now, we can prove Claim 1 as well:

[*Proof of Claim 1:* Claim 1 can be derived from Claim 2 in the same way as we derived it in the first version of the proof above. We shall not repeat this argument, since it just applies verbatim.]

But Claim 1 is simply saying that Theorem 2.70 holds under the condition that $a + b = m$. Thus, by proving Claim 1, we have shown that Theorem 2.70 holds under the condition that $a + b = m$. This completes the induction step. Thus, Theorem 2.70 is proven by strong induction. □

2.11. Induction in an interval

2.11.1. The induction principle for intervals

The induction principles we have seen so far were tailored towards proving statements whose variables range over infinite sets such as $\mathbb{N}$ and $\mathbb{Z}_{\geq g}$. Sometimes, one instead wants to do an induction on a variable that ranges over a finite interval, such as $\{g, g+1, \ldots, h\}$ for some integers g and h. We shall next state an induction principle tailored to such situations. First, we make an important convention:

Convention 2.73. If g and h are two integers such that $g > h$, then the set $\{g, g+1, \ldots, h\}$ is understood to be the empty set.

Thus, for example, $\{2, 3, \ldots, 1\} = \varnothing$ and $\{2, 3, \ldots, 0\} = \varnothing$ and $\{5, 6, \ldots, -100\} = \varnothing$. (But $\{5, 6, \ldots, 5\} = \{5\}$ and $\{5, 6, \ldots, 6\} = \{5, 6\}$.)

We now state our induction principle for intervals:

Theorem 2.74. Let $g \in \mathbb{Z}$ and $h \in \mathbb{Z}$. For each $n \in \{g, g+1, \ldots, h\}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: If $g \leq h$, then the statement $\mathcal{A}(g)$ holds.

Assumption 2: If $m \in \{g, g+1, \ldots, h-1\}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$.

Theorem 2.74 is, in a sense, the closest one can get to Theorem 2.53 when having only finitely many statements $\mathcal{A}(g), \mathcal{A}(g+1), \ldots, \mathcal{A}(h)$ instead of an infinite sequence of statements $\mathcal{A}(g), \mathcal{A}(g+1), \mathcal{A}(g+2), \ldots$. It is easy to derive Theorem 2.74 from Corollary 2.61:

Proof of Theorem 2.74. For each $n \in \mathbb{Z}_{\geq g}$, we define $\mathcal{B}(n)$ to be the logical statement

$$(\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds}).$$

Now, let us consider the Assumptions A and B from Corollary 2.61. We claim that both of these assumptions are satisfied.

Assumption 1 says that if $g \leq h$, then the statement $\mathcal{A}(g)$ holds. Thus, $\mathcal{B}(g)$ holds[71]. In other words, Assumption A is satisfied.

Next, we shall prove that Assumption B is satisfied. Indeed, let $p \in \mathbb{Z}_{\geq g}$ be such that $\mathcal{B}(p)$ holds. We shall now show that $\mathcal{B}(p+1)$ also holds.

Indeed, assume that $p+1 \in \{g, g+1, \ldots, h\}$. Thus, $p+1 \leq h$, so that $p \leq p+1 \leq h$. Combining this with $p \geq g$ (since $p \in \mathbb{Z}_{\geq g}$), we conclude that $p \in \{g, g+1, \ldots, h\}$ (since p is an integer). But we have assumed that $\mathcal{B}(p)$ holds. In other words,

$$\text{if } p \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(p) \text{ holds}$$

[71] *Proof.* Assume that $g \in \{g, g+1, \ldots, h\}$. Thus, $g \leq h$. But Assumption 1 says that if $g \leq h$, then the statement $\mathcal{A}(g)$ holds. Hence, the statement $\mathcal{A}(g)$ holds (since $g \leq h$).

Now, forget that we assumed that $g \in \{g, g+1, \ldots, h\}$. We thus have proven that if $g \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(g)$ holds. In other words, $\mathcal{B}(g)$ holds (because the statement $\mathcal{B}(g)$ is defined as (if $g \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(g)$ holds)). Qed.

(because the statement $\mathcal{B}(p)$ is defined as (if $p \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(p)$ holds)). Thus, $\mathcal{A}(p)$ holds (since we have $p \in \{g, g+1, \ldots, h\}$). Also, from $p+1 \leq h$, we obtain $p \leq h-1$. Combining this with $p \geq g$, we find $p \in \{g, g+1, \ldots, h-1\}$. Thus, we know that $p \in \{g, g+1, \ldots, h-1\}$ is such that $\mathcal{A}(p)$ holds. Hence, Assumption 2 (applied to $m = p$) shows that $\mathcal{A}(p+1)$ also holds.

Now, forget that we assumed that $p+1 \in \{g, g+1, \ldots, h\}$. We thus have proven that if $p+1 \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(p+1)$ holds. In other words, $\mathcal{B}(p+1)$ holds (since the statement $\mathcal{B}(p+1)$ is defined as (if $p+1 \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(p+1)$ holds)).

Now, forget that we fixed p. We thus have proven that if $p \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds. In other words, Assumption B is satisfied.

We now know that both Assumption A and Assumption B are satisfied. Hence, Corollary 2.61 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{Z}_{\geq g}. \tag{115}$$

Now, let $n \in \{g, g+1, \ldots, h\}$. Thus, $n \geq g$, so that $n \in \mathbb{Z}_{\geq g}$. Hence, (115) shows that $\mathcal{B}(n)$ holds. In other words,

$$\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds}$$

(since the statement $\mathcal{B}(n)$ was defined as (if $n \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(n)$ holds)). Thus, $\mathcal{A}(n)$ holds (since we have $n \in \{g, g+1, \ldots, h\}$).

Now, forget that we fixed n. We thus have shown that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$. This proves Theorem 2.74. □

Theorem 2.74 is called the *principle of induction starting at g and ending at h*, and proofs that use it are usually called *proofs by induction* or *induction proofs*. As with all the other induction principles seen so far, we don't usually explicitly cite Theorem 2.74, but instead say certain words that signal that it is being applied and that (ideally) also indicate what integers g and h and what statements $\mathcal{A}(n)$ it is being applied to[72]. However, we shall reference it explicitly in our very first example of the use of Theorem 2.74:

Proposition 2.75. Let g and h be integers such that $g \leq h$. Let $b_g, b_{g+1}, \ldots, b_h$ be any $h-g+1$ nonzero integers. Assume that $b_g \geq 0$. Assume further that

$$|b_{i+1} - b_i| \leq 1 \qquad \text{for every } i \in \{g, g+1, \ldots, h-1\}. \tag{116}$$

Then, $b_n > 0$ for each $n \in \{g, g+1, \ldots, h\}$.

Proposition 2.75 is often called the "*discrete intermediate value theorem*" or the "*discrete continuity principle*". Its intuitive meaning is that if a finite list of nonzero integers starts with a nonnegative integer, and every further entry of this list differs

[72] We will explain this in Convention 2.76 below.

from its preceding entry by at most 1, then all entries of this list must be positive. An example of such a list is $(2,3,3,2,3,4,4,3,2,3,2,3,2,1)$. Notice that Proposition 2.75 is, again, rather obvious from an intuitive perspective: It just says that it isn't possible to go from a nonnegative integer to a negative integer by steps of 1 without ever stepping at 0. The rigorous proof of Proposition 2.75 is not much harder – but because it is a statement about elements of $\{g, g+1, \ldots, h\}$, it naturally relies on Theorem 2.74:

Proof of Proposition 2.75. For each $n \in \{g, g+1, \ldots, h\}$, we let $\mathcal{A}(n)$ be the statement $(b_n > 0)$.

Our next goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \{g, g+1, \ldots, h\}$.

All the $h-g+1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_g is nonzero. In other words, $b_g \neq 0$. Combining this with $b_g \geq 0$, we obtain $b_g > 0$. In other words, the statement $\mathcal{A}(g)$ holds (since this statement $\mathcal{A}(g)$ is defined to be $(b_g > 0)$). Hence,

$$\text{if } g \leq h \text{, then the statement } \mathcal{A}(g) \text{ holds.} \tag{117}$$

Now, we claim that

$$\text{if } m \in \{g, g+1, \ldots, h-1\} \text{ is such that } \mathcal{A}(m) \text{ holds, then } \mathcal{A}(m+1) \text{ also holds.} \tag{118}$$

[*Proof of (118):* Let $m \in \{g, g+1, \ldots, h-1\}$ be such that $\mathcal{A}(m)$ holds. We must show that $\mathcal{A}(m+1)$ also holds.

We have assumed that $\mathcal{A}(m)$ holds. In other words, $b_m > 0$ holds (since $\mathcal{A}(m)$ is defined to be the statement $(b_m > 0)$). Now, (116) (applied to $i = m$) yields $|b_{m+1} - b_m| \leq 1$. But it is well-known (and easy to see) that every integer x satisfies $-x \leq |x|$. Applying this to $x = b_{m+1} - b_m$, we obtain $-(b_{m+1} - b_m) \leq |b_{m+1} - b_m| \leq 1$. In other words, $1 \geq -(b_{m+1} - b_m) = b_m - b_{m+1}$. In other words, $1 + b_{m+1} \geq b_m$. Hence, $1 + b_{m+1} \geq b_m > 0$, so that $1 + b_{m+1} \geq 1$ (since $1 + b_{m+1}$ is an integer). In other words, $b_{m+1} \geq 0$.

But all the $h-g+1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_{m+1} is nonzero. In other words, $b_{m+1} \neq 0$. Combining this with $b_{m+1} \geq 0$, we obtain $b_{m+1} > 0$. But this is precisely the statement $\mathcal{A}(m+1)$ (because $\mathcal{A}(m+1)$ is defined to be the statement $(b_{m+1} > 0)$). Thus, the statement $\mathcal{A}(m+1)$ holds.

Now, forget that we fixed m. We thus have shown that if $m \in \{g, g+1, \ldots, h-1\}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. This proves (118).]

Now, both assumptions of Theorem 2.74 are satisfied (indeed, Assumption 1 holds because of (117), whereas Assumption 2 holds because of (118)). Thus, Theorem 2.74 shows that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$. In other words, $b_n > 0$ holds for each $n \in \{g, g+1, \ldots, h\}$ (since $\mathcal{A}(n)$ is the statement $(b_n > 0)$). This proves Proposition 2.75. $\square$

2.11.2. Conventions for writing induction proofs in intervals

Next, we shall introduce some standard language that is commonly used in proofs by induction starting at g and ending at h. This language closely imitates the one we use for proofs by standard induction:

Convention 2.76. Let $g \in \mathbb{Z}$ and $h \in \mathbb{Z}$. For each $n \in \{g, g+1, \ldots, h\}$, let $\mathcal{A}(n)$ be a logical statement. Assume that you want to prove that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$.

Theorem 2.74 offers the following strategy for proving this: First show that Assumption 1 of Theorem 2.74 is satisfied; then, show that Assumption 2 of Theorem 2.74 is satisfied; then, Theorem 2.74 automatically completes your proof.

A proof that follows this strategy is called a *proof by induction on n* (or *proof by induction over n*) *starting at g and ending at h* or (less precisely) an *inductive proof*. Most of the time, the words "starting at g and ending at h" are omitted, since they merely repeat what is clear from the context anyway: For example, if you make a claim about all integers $n \in \{3, 4, 5, 6\}$, and you say that you are proving it by induction on n, it is clear that you are using induction on n starting at 3 and ending at 6.

The proof that Assumption 1 is satisfied is called the *induction base* (or *base case*) of the proof. The proof that Assumption 2 is satisfied is called the *induction step* of the proof.

In order to prove that Assumption 2 is satisfied, you will usually want to fix an $m \in \{g, g+1, \ldots, h-1\}$ such that $\mathcal{A}(m)$ holds, and then prove that $\mathcal{A}(m+1)$ holds. In other words, you will usually want to fix $m \in \{g, g+1, \ldots, h-1\}$, assume that $\mathcal{A}(m)$ holds, and then prove that $\mathcal{A}(m+1)$ holds. When doing so, it is common to refer to the assumption that $\mathcal{A}(m)$ holds as the *induction hypothesis* (or *induction assumption*).

Unsurprisingly, this language parallels the language introduced in Convention 2.3 and in Convention 2.56.

Again, we can shorten our inductive proofs by omitting some sentences that convey no information. In particular, we can leave out the explicit definition of the statement $\mathcal{A}(n)$ when this statement is precisely the claim that we are proving (without the "for each $n \in \{g, g+1, \ldots, h\}$" part). Furthermore, it is common to leave the "If $g \leq h$" part of Assumption 1 unsaid (i.e., to pretend that Assumption 1 simply says that $\mathcal{A}(g)$ holds). Strictly speaking, this is somewhat imprecise, since $\mathcal{A}(g)$ is not defined when $g > h$; but of course, the whole claim that is being proven is moot anyway when $g > h$ (because there exist no $n \in \{g, g+1, \ldots, h\}$ in this case), so this imprecision doesn't matter.

Thus, we can rewrite our above proof of Proposition 2.75 as follows:

Proof of Proposition 2.75 (second version). We claim that

$$b_n > 0 \tag{119}$$

for each $n \in \{g, g+1, \ldots, h\}$.

Indeed, we shall prove (119) by induction on n:

Induction base: All the $h - g + 1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_g is nonzero. In other words, $b_g \neq 0$. Combining this with $b_g \geq 0$, we obtain $b_g > 0$. In other words, (119) holds for $n = g$. This completes the induction base.

Induction step: Let $m \in \{g, g+1, \ldots, h-1\}$. Assume that (119) holds for $n = m$. We must show that (119) also holds for $n = m + 1$.

We have assumed that (119) holds for $n = m$. In other words, $b_m > 0$. Now, (116) (applied to $i = m$) yields $|b_{m+1} - b_m| \leq 1$. But it is well-known (and easy to see) that every integer x satisfies $-x \leq |x|$. Applying this to $x = b_{m+1} - b_m$, we obtain $-(b_{m+1} - b_m) \leq |b_{m+1} - b_m| \leq 1$. In other words, $1 \geq -(b_{m+1} - b_m) = b_m - b_{m+1}$. In other words, $1 + b_{m+1} \geq b_m$. Hence, $1 + b_{m+1} \geq b_m > 0$, so that $1 + b_{m+1} \geq 1$ (since $1 + b_{m+1}$ is an integer). In other words, $b_{m+1} \geq 0$.

But all the $h - g + 1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_{m+1} is nonzero. In other words, $b_{m+1} \neq 0$. Combining this with $b_{m+1} \geq 0$, we obtain $b_{m+1} > 0$. In other words, (119) holds for $n = m + 1$. This completes the induction step. Thus, (119) is proven by induction. This proves Proposition 2.75. □

2.12. Strong induction in an interval

2.12.1. The strong induction principle for intervals

We shall next state yet another induction principle – one that combines the idea of strong induction (as in Theorem 2.60) with the idea of working inside an interval $\{g, g+1, \ldots, h\}$ (as in Theorem 2.74):

> **Theorem 2.77.** Let $g \in \mathbb{Z}$ and $h \in \mathbb{Z}$. For each $n \in \{g, g+1, \ldots, h\}$, let $\mathcal{A}(n)$ be a logical statement.
>
> Assume the following:
>
> *Assumption 1:* If $m \in \{g, g+1, \ldots, h\}$ is such that
>
> $$(\mathcal{A}(n) \text{ holds for every } n \in \{g, g+1, \ldots, h\} \text{ satisfying } n < m),$$
>
> then $\mathcal{A}(m)$ holds.
>
> Then, $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$.

Our proof of Theorem 2.77 will be similar to the proof of Theorem 2.74, except that we shall be using Theorem 2.60 instead of Corollary 2.61. Or, to be more precise, we shall be using the following restatement of Theorem 2.60:

Corollary 2.78. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{B}(n)$ be a logical statement.

Assume the following:

Assumption A: If $p \in \mathbb{Z}_{\geq g}$ is such that

$$\left(\mathcal{B}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < p\right),$$

then $\mathcal{B}(p)$ holds.

Then, $\mathcal{B}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Proof of Corollary 2.78. Corollary 2.78 is exactly Theorem 2.60, except that some names have been changed:

- The statements $\mathcal{A}(n)$ have been renamed as $\mathcal{B}(n)$.
- Assumption 1 has been renamed as Assumption A.
- The variable m in Assumption A has been renamed as p.

Thus, Corollary 2.78 holds (since Theorem 2.60 holds). □

We can now prove Theorem 2.77:

Proof of Theorem 2.77. For each $n \in \mathbb{Z}_{\geq g}$, we define $\mathcal{B}(n)$ to be the logical statement

$$(\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds}).$$

Now, let us consider the Assumption A from Corollary 2.78. We claim that this assumption is satisfied.

Indeed, let $p \in \mathbb{Z}_{\geq g}$ be such that

$$\left(\mathcal{B}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < p\right). \tag{120}$$

We shall now show that $\mathcal{B}(p)$ holds.

Indeed, assume that $p \in \{g, g+1, \ldots, h\}$. Thus, $p \leq h$.

Now, let $n \in \{g, g+1, \ldots, h\}$ be such that $n < p$. Then, $n \in \{g, g+1, \ldots, h\} \subseteq \{g, g+1, g+2, \ldots\} = \mathbb{Z}_{\geq g}$ and $n < p$. Hence, (120) shows that $\mathcal{B}(n)$ holds. In other words, $(\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds})$ (because the statement $\mathcal{B}(n)$ is defined as $(\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds})$). Therefore, $\mathcal{A}(n)$ holds (since we know that $n \in \{g, g+1, \ldots, h\}$).

Now, forget that we fixed n. We thus have proven that

$$(\mathcal{A}(n) \text{ holds for every } n \in \{g, g+1, \ldots, h\} \text{ satisfying } n < p).$$

Hence, Assumption 1 (applied to $m = p$) yields that $\mathcal{A}(p)$ holds.

Now, forget that we assumed that $p \in \{g, g+1, \ldots, h\}$. We thus have proven that

$$(\text{if } p \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(p) \text{ holds}).$$

In other words, $\mathcal{B}(p)$ holds (since the statement $\mathcal{B}(p)$ was defined as $(\text{if } p \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(p) \text{ holds})$).

Now, forget that we fixed p. We thus have shown that if $p \in \mathbb{Z}_{\geq g}$ is such that

$$(\mathcal{B}(n) \text{ holds for every } n \in \mathbb{Z}_{\geq g} \text{ satisfying } n < p),$$

then $\mathcal{B}(p)$ holds. In other words, Assumption A is satisfied.

Hence, Corollary 2.78 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{Z}_{\geq g}. \tag{121}$$

Now, let $n \in \{g, g+1, \ldots, h\}$. Thus, $n \geq g$, so that $n \in \mathbb{Z}_{\geq g}$. Hence, (121) shows that $\mathcal{B}(n)$ holds. In other words,

$$\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds}$$

(since the statement $\mathcal{B}(n)$ was defined as $(\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds})$). Thus, $\mathcal{A}(n)$ holds (since we have $n \in \{g, g+1, \ldots, h\}$).

Now, forget that we fixed n. We thus have shown that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$. This proves Theorem 2.77. □

Theorem 2.77 is called the *principle of strong induction starting at g and ending at h*, and proofs that use it are usually called *proofs by strong induction*. Once again, we usually don't explicitly cite Theorem 2.77 in such proofs, and we usually don't say explicitly what g and h are and what the statements $\mathcal{A}(n)$ are when it is clear from the context. But (as with all the other induction principles considered so far) we shall be explicit about all these details in our first example:

Proposition 2.79. Let g and h be integers such that $g \leq h$. Let $b_g, b_{g+1}, \ldots, b_h$ be any $h - g + 1$ nonzero integers. Assume that $b_g \geq 0$. Assume that for each $p \in \{g+1, g+2, \ldots, h\}$,

$$\text{there exists some } j \in \{g, g+1, \ldots, p-1\} \text{ such that } b_p \geq b_j - 1. \tag{122}$$

(Of course, the j can depend on p.) Then, $b_n > 0$ for each $n \in \{g, g+1, \ldots, h\}$.

Proposition 2.79 is a more general (although less intuitive) version of Proposition 2.75; indeed, it is easy to see that the condition (116) is stronger than the condition (122) (when required for all $p \in \{g+1, g+2, \ldots, h\}$).

Example 2.80. For this example, set $g = 3$ and $h = 7$. Then, if we set $(b_3, b_4, b_5, b_6, b_7) = (4, 5, 3, 4, 2)$, then the condition (122) holds for all $p \in \{g+1, g+2, \ldots, h\}$. (For example, it holds for $p = 5$, since $b_5 = 3 \geq 4 - 1 = b_1 - 1$ and $1 \in \{g, g+1, \ldots, 5-1\}$.) On the other hand, if we set $(b_3, b_4, b_5, b_6, b_7) = (4, 5, 2, 4, 3)$, then this condition does not hold (indeed, it fails for $p = 5$, since $b_5 = 2$ is neither $\geq 4 - 1$ nor $\geq 5 - 1$).

Let us now prove Proposition 2.79 using Theorem 2.77:

Proof of Proposition 2.79. For each $n \in \{g, g+1, \ldots, h\}$, we let $\mathcal{A}(n)$ be the statement $(b_n > 0)$.

Our next goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \{g, g+1, \ldots, h\}$.

All the $h - g + 1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_g is nonzero. In other words, $b_g \neq 0$. Combining this with $b_g \geq 0$, we obtain $b_g > 0$. In other words, the statement $\mathcal{A}(g)$ holds (since this statement $\mathcal{A}(g)$ is defined to be $(b_g > 0)$).

Now, we make the following claim:

Claim 1: If $m \in \{g, g+1, \ldots, h\}$ is such that

$$(\mathcal{A}(n) \text{ holds for every } n \in \{g, g+1, \ldots, h\} \text{ satisfying } n < m),$$

then $\mathcal{A}(m)$ holds.

[*Proof of Claim 1:* Let $m \in \{g, g+1, \ldots, h\}$ be such that

$$(\mathcal{A}(n) \text{ holds for every } n \in \{g, g+1, \ldots, h\} \text{ satisfying } n < m). \tag{123}$$

We must show that $\mathcal{A}(m)$ holds.

If $m = g$, then this follows from the fact that $\mathcal{A}(g)$ holds. Thus, for the rest of the proof of Claim 1, we WLOG assume that we don't have $m = g$. Hence, $m \neq g$. Combining this with $m \in \{g, g+1, \ldots, h\}$, we obtain $m \in \{g, g+1, \ldots, h\} \setminus \{g\} \subseteq \{g+1, g+2, \ldots, h\}$. Hence, (122) (applied to $p = m$) shows that there exists some $j \in \{g, g+1, \ldots, m-1\}$ such that $b_m \geq b_j - 1$. Consider this j. From $m \in \{g+1, g+2, \ldots, h\}$, we obtain $m \leq h$.

From $j \in \{g, g+1, \ldots, m-1\}$, we obtain $j \leq m - 1 < m$. Also, $j \in \{g, g+1, \ldots, m-1\} \subseteq \{g, g+1, \ldots, h\}$ (since $m - 1 \leq m \leq h$). Thus, (123) (applied to $n = j$) yields that $\mathcal{A}(j)$ holds. In other words, $b_j > 0$ holds (since $\mathcal{A}(j)$ is defined to be the statement $(b_j > 0)$). Thus, $b_j \geq 1$ (since b_j is an integer), so that $b_j - 1 \geq 0$. But recall that $b_m \geq b_j - 1 \geq 0$.

But all the $h - g + 1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_m is nonzero. In other words, $b_m \neq 0$. Combining this with $b_m \geq 0$, we obtain $b_m > 0$. But this is precisely the statement $\mathcal{A}(m)$ (because $\mathcal{A}(m)$ is defined to be the statement $(b_m > 0)$). Thus, the statement $\mathcal{A}(m)$ holds. This completes the proof of Claim 1.]

Claim 1 says that Assumption 1 of Theorem 2.77 is satisfied. Thus, Theorem 2.77 shows that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$. In other words, $b_n > 0$ holds for each $n \in \{g, g+1, \ldots, h\}$ (since $\mathcal{A}(n)$ is the statement $(b_n > 0)$). This proves Proposition 2.79. $\square$

2.12.2. Conventions for writing strong induction proofs in intervals

Next, we shall introduce some standard language that is commonly used in proofs by strong induction starting at g and ending at h. This language closely imitates the one we use for proofs by "usual" strong induction:

Convention 2.81. Let $g \in \mathbb{Z}$ and $h \in \mathbb{Z}$. For each $n \in \{g, g+1, \ldots, h\}$, let $\mathcal{A}(n)$ be a logical statement. Assume that you want to prove that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$.

Theorem 2.77 offers the following strategy for proving this: Show that Assumption 1 of Theorem 2.77 is satisfied; then, Theorem 2.77 automatically completes your proof.

A proof that follows this strategy is called a *proof by strong induction on n starting at g and ending at h*. Most of the time, the words "starting at g and ending at h" are omitted. The proof that Assumption 1 is satisfied is called the *induction step* of the proof. This kind of proof has no "induction base".

In order to prove that Assumption 1 is satisfied, you will usually want to fix an $m \in \{g, g+1, \ldots, h\}$ such that

$$(\mathcal{A}(n) \text{ holds for every } n \in \{g, g+1, \ldots, h\} \text{ satisfying } n < m), \tag{124}$$

and then prove that $\mathcal{A}(m)$ holds. In other words, you will usually want to fix $m \in \{g, g+1, \ldots, h\}$, assume that (124) holds, and then prove that $\mathcal{A}(m)$ holds. When doing so, it is common to refer to the assumption that (124) holds as the *induction hypothesis* (or *induction assumption*).

Unsurprisingly, this language parallels the language introduced in Convention 2.63.

As before, proofs using strong induction can be shortened by leaving out some uninformative prose. In particular, the explicit definition of the statement $\mathcal{A}(n)$ can often be omitted when this statement is precisely the claim that we are proving (without the "for each $n \in \{g, g+1, \ldots, h\}$" part). The values of g and h can also be inferred from the statement of the claim, so they don't need to be specified explicitly. And once again, we don't need to write "*Induction step:*", since our strong induction has no induction base.

This leads to the following abridged version of our above proof of Proposition 2.79:

Proof of Proposition 2.79 (second version). We claim that

$$b_n > 0 \tag{125}$$

for each $n \in \{g, g+1, \ldots, h\}$.

Indeed, we shall prove (125) by strong induction on n:

Let $m \in \{g, g+1, \ldots, h\}$. Assume that (125) holds for every $n \in \{g, g+1, \ldots, h\}$ satisfying $n < m$. We must show that (125) also holds for $n = m$. In other words, we must show that $b_m > 0$.

All the $h - g + 1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_g is nonzero. In other words, $b_g \neq 0$. Combining this with $b_g \geq 0$, we obtain $b_g > 0$.

We have assumed that (125) holds for every $n \in \{g, g+1, \ldots, h\}$ satisfying $n < m$. In other words, we have

$$b_n > 0 \text{ for every } n \in \{g, g+1, \ldots, h\} \text{ satisfying } n < m. \tag{126}$$

Recall that we must prove that $b_m > 0$. If $m = g$, then this follows from $b_g > 0$. Thus, for the rest of this induction step, we WLOG assume that we don't have $m = g$. Hence, $m \neq g$. Combining this with $m \in \{g, g+1, \ldots, h\}$, we obtain $m \in \{g, g+1, \ldots, h\} \setminus \{g\} \subseteq \{g+1, g+2, \ldots, h\}$. Hence, (122) (applied to $p = m$) shows that there exists some $j \in \{g, g+1, \ldots, m-1\}$ such that $b_m \geq b_j - 1$. Consider this j. From $m \in \{g+1, g+2, \ldots, h\}$, we obtain $m \leq h$.

From $j \in \{g, g+1, \ldots, m-1\}$, we obtain $j \leq m - 1 < m$. Also, $j \in \{g, g+1, \ldots, m-1\} \subseteq \{g, g+1, \ldots, h\}$ (since $m - 1 \leq m \leq h$). Thus, (126) (applied to $n = j$) yields that $b_j > 0$. Thus, $b_j \geq 1$ (since b_j is an integer), so that $b_j - 1 \geq 0$. But recall that $b_m \geq b_j - 1 \geq 0$.

But all the $h - g + 1$ integers $b_g, b_{g+1}, \ldots, b_h$ are nonzero (by assumption). Thus, in particular, b_m is nonzero. In other words, $b_m \neq 0$. Combining this with $b_m \geq 0$, we obtain $b_m > 0$.

Thus, we have proven that $b_m > 0$. In other words, (125) holds for $n = m$. This completes the induction step. Thus, (125) is proven by strong induction. This proves Proposition 2.79. □

2.13. General associativity for composition of maps

2.13.1. Associativity of map composition

Recall that if $f : X \to Y$ and $g : Y \to Z$ are two maps, then the *composition* $g \circ f$ of the maps g and f is defined to be the map

$$X \to Z, \ x \mapsto g(f(x)).$$

Now, if we have four sets X, Y, Z and W and three maps $c : X \to Y$, $b : Y \to Z$ and $a : Z \to W$, then we can build two possible compositions that use all three of these maps: namely, the two compositions $(a \circ b) \circ c$ and $a \circ (b \circ c)$. It turns out that these two compositions are the same map:[73]

[73]Of course, when some of the four sets X, Y, Z and W are equal, then more compositions can be built: For example, if $Y = Z = W$, then we can also build the composition $(b \circ a) \circ c$ or the composition $((b \circ b) \circ a) \circ c$. But these compositions are not the same map as the two that we previously constructed.

Proposition 2.82. Let X, Y, Z and W be four sets. Let $c : X \to Y$, $b : Y \to Z$ and $a : Z \to W$ be three maps. Then,

$$(a \circ b) \circ c = a \circ (b \circ c).$$

Proposition 2.82 is called the *associativity of map composition*, and is proven straightforwardly:

Proof of Proposition 2.82. Let $x \in X$. Then, the definition of $b \circ c$ yields $(b \circ c)(x) = b(c(x))$. But the definition of $(a \circ b) \circ c$ yields

$$((a \circ b) \circ c)(x) = (a \circ b)(c(x)) = a(b(c(x))) \qquad \text{(by the definition of } a \circ b\text{)}.$$

On the other hand, the definition of $a \circ (b \circ c)$ yields

$$(a \circ (b \circ c))(x) = a\left(\underbrace{(b \circ c)(x)}_{=b(c(x))}\right) = a(b(c(x))).$$

Comparing these two equalities, we obtain $((a \circ b) \circ c)(x) = (a \circ (b \circ c))(x)$.

Now, forget that we fixed x. We thus have shown that

$$((a \circ b) \circ c)(x) = (a \circ (b \circ c))(x) \qquad \text{for each } x \in X.$$

In other words, $(a \circ b) \circ c = a \circ (b \circ c)$. This proves Proposition 2.82. □

2.13.2. Composing more than 3 maps: exploration

Proposition 2.82 can be restated as follows: If a, b and c are three maps such that the compositions $a \circ b$ and $b \circ c$ are well-defined, then $(a \circ b) \circ c = a \circ (b \circ c)$. This allows us to write "$a \circ b \circ c$" for each of the compositions $(a \circ b) \circ c$ and $a \circ (b \circ c)$ without having to disambiguate this expression by means of parentheses. It is natural to ask whether we can do the same thing for more than three maps. For example, let us consider four maps a, b, c and d for which the compositions $a \circ b$, $b \circ c$ and $c \circ d$ are well-defined:

Example 2.83. Let X, Y, Z, W and U be five sets. Let $d : X \to Y$, $c : Y \to Z$, $b : Z \to W$ and $a : W \to U$ be four maps. Then, there we can construct five compositions that use all four of these maps; these five compositions are

$$((a \circ b) \circ c) \circ d, \qquad (a \circ (b \circ c)) \circ d, \qquad (a \circ b) \circ (c \circ d), \tag{127}$$

$$a \circ ((b \circ c) \circ d), \qquad a \circ (b \circ (c \circ d)). \tag{128}$$

It turns out that these five compositions are all the same map. Indeed, this follows by combining the following observations:

- We have $((a \circ b) \circ c) \circ d = (a \circ (b \circ c)) \circ d$ (since Proposition 2.82 yields $(a \circ b) \circ c = a \circ (b \circ c)$).
- We have $a \circ ((b \circ c) \circ d) = a \circ (b \circ (c \circ d))$ (since Proposition 2.82 yields $(b \circ c) \circ d = b \circ (c \circ d)$).
- We have $(a \circ (b \circ c)) \circ d = a \circ ((b \circ c) \circ d)$ (by Proposition 2.82, applied to W, U, $b \circ c$ and d instead of Z, W, b and c).
- We have $((a \circ b) \circ c) \circ d = (a \circ b) \circ (c \circ d)$ (by Proposition 2.82, applied to U, $a \circ b$, c and d instead of W, a, b and c).

Hence, all five compositions are equal. Thus, we can write "$a \circ b \circ c \circ d$" for each of these five compositions, again dropping the parentheses.

We shall refer to the five compositions listed in (127) and (128) as the "complete parenthesizations of $a \circ b \circ c \circ d$". Here, the word "parenthesization" means a way to put parentheses into the expression "$a \circ b \circ c \circ d$", whereas the word "complete" means that these parentheses unambiguously determine which two maps any given $\circ$ sign is composing. (For example, the parenthesization "$(a \circ b \circ c) \circ d$" is not complete, because the first $\circ$ sign in it could be either composing a with b or composing a with $b \circ c$. But the parenthesization "$((a \circ b) \circ c) \circ d$" is complete, because its first $\circ$ sign composes a and b, whereas its second $\circ$ sign composes $a \circ b$ with c, and finally its third $\circ$ sign composes $(a \circ b) \circ c$ with d.)

Thus, we have seen that all five complete parenthesizations of $a \circ b \circ c \circ d$ are the same map.

What happens if we compose more than four maps? Clearly, the more maps we have, the more complete parenthesizations can be constructed. We have good reasons to suspect that these parenthesizations will all be the same map (so we can again drop the parentheses); but if we try to prove it in the ad-hoc way we did in Example 2.83, then we have more and more work to do the more maps we are composing. Clearly, if we want to prove our suspicion for arbitrarily many maps, we need a more general approach.

2.13.3. Formalizing general associativity

So let us make a general statement; but first, let us formally define the notion of a "complete parenthesization":

Definition 2.84. Let n be a positive integer. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then, we want to define the notion of a *complete parenthesization* of $f_n \circ f_{n-1} \circ \cdots \circ f_1$. We define this notion by recursion on n as follows:

- For $n = 1$, there is only one complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$, and this is simply the map $f_1 : X_1 \to X_2$.
- If $n > 1$, then the complete parenthesizations of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ are all the maps of the form $\alpha \circ \beta$, where
  - k is some element of $\{1, 2, \ldots, n-1\}$;
  - α is a complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_{k+1}$;
  - β is a complete parenthesization of $f_k \circ f_{k-1} \circ \cdots \circ f_1$.

Example 2.85. Let us see what this definition yields for small values of n:

- For $n = 1$, the only complete parenthesization of f_1 is f_1.
- For $n = 2$, the only complete parenthesization of $f_2 \circ f_1$ is the composition $f_2 \circ f_1$ (because here, the only possible values of k, α and β are 1, f_2 and f_1, respectively).
- For $n = 3$, the complete parenthesizations of $f_3 \circ f_2 \circ f_1$ are the two compositions $(f_3 \circ f_2) \circ f_1$ and $f_3 \circ (f_2 \circ f_1)$ (because here, the only possible values of k are 1 and 2, and each value of k uniquely determines α and β). Proposition 2.82 shows that they are equal (as maps).
- For $n = 4$, the complete parenthesizations of $f_4 \circ f_3 \circ f_2 \circ f_1$ are the five compositions

  $$((f_4 \circ f_3) \circ f_2) \circ f_1, \qquad (f_4 \circ (f_3 \circ f_2)) \circ f_1, \qquad (f_4 \circ f_3) \circ (f_2 \circ f_1),$$
  $$f_4 \circ ((f_3 \circ f_2) \circ f_1), \qquad f_4 \circ (f_3 \circ (f_2 \circ f_1)).$$

  (These are exactly the five compositions listed in (127) and (128), except that the maps d, c, b, a are now called f_1, f_2, f_3, f_4.) We have seen in Example 2.83 that these five compositions are equal as maps.
- For $n = 5$, the complete parenthesizations of $f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1$ are 14 compositions, one of which is $(f_5 \circ f_4) \circ (f_3 \circ (f_2 \circ f_1))$. Again, it is laborious but not difficult to check that all the 14 compositions are equal as maps.

Now, we want to prove the following general statement:

Theorem 2.86. Let n be a positive integer. Let $X_1, X_2, \ldots, X_{n+1}$ be $n + 1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then, all complete parenthesizations of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ are the same map (from X_1 to X_{n+1}).

Theorem 2.86 is sometimes called the *general associativity* theorem, and is often proved in the context of monoids (see, e.g., [Artin10, Proposition 2.1.4]); while the context is somewhat different from ours, the proofs usually given still apply in ours.

2.13.4. Defining the "canonical" composition $C(f_n, f_{n-1}, \ldots, f_1)$

We shall prove Theorem 2.86 in a slightly indirect way: We first define a *specific* complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$, which we shall call $C(f_n, f_{n-1}, \ldots, f_1)$; then we will show that it satisfies certain equalities (Proposition 2.89), and then prove that every complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ equals this map $C(f_n, f_{n-1}, \ldots, f_1)$ (Proposition 2.90). Each step of this strategy will rely on induction.

We begin with the definition of $C(f_n, f_{n-1}, \ldots, f_1)$:

Definition 2.87. Let n be a positive integer. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then, we want to define a map $C(f_n, f_{n-1}, \ldots, f_1) : X_1 \to X_{n+1}$. We define this map by recursion on n as follows:

- If $n = 1$, then we define $C(f_n, f_{n-1}, \ldots, f_1)$ to be the map $f_1 : X_1 \to X_2$. (Note that in this case, $C(f_n, f_{n-1}, \ldots, f_1) = C(f_1)$, because $(f_n, f_{n-1}, \ldots, f_1) = (f_1, f_{1-1}, \ldots, f_1) = (f_1)$.)
- If $n > 1$, then we define $C(f_n, f_{n-1}, \ldots, f_1) : X_1 \to X_{n+1}$ by

$$C(f_n, f_{n-1}, \ldots, f_1) = f_n \circ C(f_{n-1}, f_{n-2}, \ldots, f_1). \tag{129}$$

Example 2.88. Consider the situation of Definition 2.87.

(a) If $n = 1$, then

$$C(f_1) = f_1 \tag{130}$$

(by the $n = 1$ case of the definition).

(b) If $n = 2$, then

$$C(f_2, f_1) = f_2 \circ \underbrace{C(f_1)}_{\substack{= f_1 \\ \text{(by (130))}}} \qquad \text{(by (129), applied to } n = 2\text{)}$$
$$= f_2 \circ f_1. \tag{131}$$

(c) If $n = 3$, then

$$C(f_3, f_2, f_1) = f_3 \circ \underbrace{C(f_2, f_1)}_{\substack{= f_2 \circ f_1 \\ \text{(by (131))}}} \qquad \text{(by (129), applied to } n = 3\text{)}$$
$$= f_3 \circ (f_2 \circ f_1). \tag{132}$$

(d) If $n = 4$, then

$$C(f_4, f_3, f_2, f_1) = f_4 \circ \underbrace{C(f_3, f_2, f_1)}_{\substack{= f_3 \circ (f_2 \circ f_1) \\ \text{(by (132))}}} \qquad \text{(by (129), applied to } n = 4\text{)}$$
$$= f_4 \circ (f_3 \circ (f_2 \circ f_1)). \tag{133}$$

(e) For an arbitrary $n \geq 1$, we can informally write $C(f_n, f_{n-1}, \ldots, f_1)$ as

$$C(f_n, f_{n-1}, \ldots, f_1) = f_n \circ (f_{n-1} \circ (f_{n-2} \circ (\cdots \circ (f_2 \circ f_1) \cdots))).$$

The right hand side of this equality is a complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$, where all the parentheses are "concentrated as far right as possible" (i.e., there is an opening parenthesis after each "$\circ$" sign except for the last one; and there are $n-2$ closing parentheses at the end of the expression). This is merely a visual restatement of the recursive definition of $C(f_n, f_{n-1}, \ldots, f_1)$ we gave above.

2.13.5. The crucial property of $C(f_n, f_{n-1}, \ldots, f_1)$

The following proposition will be key to our proof of Theorem 2.86:

Proposition 2.89. Let n be a positive integer. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then,

$$C(f_n, f_{n-1}, \ldots, f_1) = C(f_n, f_{n-1}, \ldots, f_{k+1}) \circ C(f_k, f_{k-1}, \ldots, f_1)$$

for each $k \in \{1, 2, \ldots, n-1\}$.

Proof of Proposition 2.89. Forget that we fixed n, $X_1, X_2, \ldots, X_{n+1}$ and the maps f_i. We shall prove Proposition 2.89 by induction on n: [74]

Induction base: If $n = 1$, then $\{1, 2, \ldots, n-1\} = \{1, 2, \ldots, 1-1\} = \varnothing$. Hence, if $n = 1$, then there exists no $k \in \{1, 2, \ldots, n-1\}$. Thus, if $n = 1$, then Proposition 2.89 is vacuously true (since Proposition 2.89 has a "for each $k \in \{1, 2, \ldots, n-1\}$" clause). This completes the induction base.

Induction step: Let $m \in \mathbb{Z}_{\geq 1}$. Assume that Proposition 2.89 holds under the condition that $n = m$. We must now prove that Proposition 2.89 holds under the condition that $n = m+1$. In other words, we must prove the following claim:

Claim 1: Let $X_1, X_2, \ldots, X_{(m+1)+1}$ be $(m+1)+1$ sets. For each $i \in \{1, 2, \ldots, m+1\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then,

$$C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_1\right) = C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right) \circ C(f_k, f_{k-1}, \ldots, f_1)$$

[74] The induction principle that we are applying here is Theorem 2.53 with $g = 1$ (since $\mathbb{Z}_{\geq 1}$ is the set of all positive integers).

for each $k \in \{1, 2, \ldots, (m+1) - 1\}$.

[*Proof of Claim 1:* Let $k \in \{1, 2, \ldots, (m+1) - 1\}$. Thus, $k \in \{1, 2, \ldots, (m+1) - 1\} = \{1, 2, \ldots, m\}$ (since $(m+1) - 1 = m$).

We know that $X_1, X_2, \ldots, X_{(m+1)+1}$ are $(m+1) + 1$ sets. In other words, $X_1, X_2, \ldots, X_{m+2}$ are $m+2$ sets (since $(m+1) + 1 = m+2$). We have $m \in \mathbb{Z}_{\geq 1}$, thus $m \geq 1 > 0$; hence, $m + 1 > 1$. Thus, (129) (applied to $n = m+1$) yields

$$\begin{aligned} C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_1\right) &= f_{m+1} \circ C\left(f_{(m+1)-1}, f_{(m+1)-2}, \ldots, f_1\right) \\ &= f_{m+1} \circ C\left(f_m, f_{m-1}, \ldots, f_1\right) \end{aligned} \tag{134}$$

(since $(m+1) - 1 = m$ and $(m+1) - 2 = m - 1$).

But we are in one of the following two cases:

Case 1: We have $k = m$.

Case 2: We have $k \neq m$.

Let us first consider Case 1. In this case, we have $k = m$. Hence,

$$C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right) = C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{m+1}\right) = C\left(f_{m+1}\right) = f_{m+1}$$

(by (130), applied to X_{m+1}, X_{m+2} and f_{m+1} instead of X_1, X_2 and f_1), so that

$$\underbrace{C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right)}_{= f_{m+1}} \circ \underbrace{C\left(f_k, f_{k-1}, \ldots, f_1\right)}_{\substack{= C(f_m, f_{m-1}, \ldots, f_1) \\ (\text{since } k = m)}} = f_{m+1} \circ C\left(f_m, f_{m-1}, \ldots, f_1\right).$$

Comparing this with (134), we obtain

$$C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_1\right) = C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right) \circ C\left(f_k, f_{k-1}, \ldots, f_1\right).$$

Hence, Claim 1 is proven in Case 1.

Let us now consider Case 2. In this case, we have $k \neq m$. Combining $k \in \{1, 2, \ldots, m\}$ with $k \neq m$, we obtain

$$k \in \{1, 2, \ldots, m\} \setminus \{m\} = \{1, 2, \ldots, m-1\}.$$

Hence, $k \leq m - 1 < m$, so that $m + 1 - \underbrace{k}_{< m} > m + 1 - m = 1$.

But we assumed that Proposition 2.89 holds under the condition that $n = m$. Hence, we can apply Proposition 2.89 to m instead of n. We thus obtain

$$C\left(f_m, f_{m-1}, \ldots, f_1\right) = C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right) \circ C\left(f_k, f_{k-1}, \ldots, f_1\right)$$

(since $k \in \{1, 2, \ldots, m-1\}$). Now, (134) yields

$$\begin{aligned} & C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_1\right) \\ &= f_{m+1} \circ \underbrace{C\left(f_m, f_{m-1}, \ldots, f_1\right)}_{= C(f_m, f_{m-1}, \ldots, f_{k+1}) \circ C(f_k, f_{k-1}, \ldots, f_1)} \\ &= f_{m+1} \circ \left(C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right) \circ C\left(f_k, f_{k-1}, \ldots, f_1\right)\right). \end{aligned} \tag{135}$$

On the other hand, $m+1-k > 1$. Hence, (129) (applied to $m+1-k$, X_{k+i} and f_{k+i} instead of n, X_i and f_i) yields

$$\begin{aligned} & C\left(f_{k+(m+1-k)}, f_{k+((m+1-k)-1)}, \ldots, f_{k+1}\right) \\ & = f_{k+(m+1-k)} \circ C\left(f_{k+((m+1-k)-1)}, f_{k+((m+1-k)-2)}, \ldots, f_{k+1}\right) \\ & = f_{m+1} \circ C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right) \\ & \qquad \left(\begin{array}{c} \text{since } k+(m+1-k) = m+1 \text{ and } k+((m+1-k)-1) = m \\ \text{and } k+((m+1-k)-2) = m-1 \end{array}\right). \end{aligned}$$

Since $k+(m+1-k) = m+1$ and $k+((m+1-k)-1) = (m+1)-1$, this rewrites as

$$C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right) = f_{m+1} \circ C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right).$$

Hence,

$$\begin{aligned} & \underbrace{C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right)}_{= f_{m+1} \circ C(f_m, f_{m-1}, \ldots, f_{k+1})} \circ C\left(f_k, f_{k-1}, \ldots, f_1\right) \\ & = \left(f_{m+1} \circ C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right)\right) \circ C\left(f_k, f_{k-1}, \ldots, f_1\right) \\ & = f_{m+1} \circ \left(C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right) \circ C\left(f_k, f_{k-1}, \ldots, f_1\right)\right) \end{aligned}$$

(by Proposition 2.82, applied to $X = X_1$, $Y = X_{k+1}$, $Z = X_{m+1}$, $W = X_{m+2}$, $c = C\left(f_k, f_{k-1}, \ldots, f_1\right)$, $b = C\left(f_m, f_{m-1}, \ldots, f_{k+1}\right)$ and $a = f_{m+1}$). Comparing this with (135), we obtain

$$C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_1\right) = C\left(f_{m+1}, f_{(m+1)-1}, \ldots, f_{k+1}\right) \circ C\left(f_k, f_{k-1}, \ldots, f_1\right).$$

Hence, Claim 1 is proven in Case 2.

We have now proven Claim 1 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Claim 1 always holds.]

Now, we have proven Claim 1. In other words, we have proven that Proposition 2.89 holds under the condition that $n = m+1$. This completes the induction step. Hence, Proposition 2.89 is proven by induction. □

2.13.6. Proof of general associativity

Proposition 2.90. Let n be a positive integer. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then, every complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ equals $C\left(f_n, f_{n-1}, \ldots, f_1\right)$.

Proof of Proposition 2.90. Forget that we fixed n, $X_1, X_2, \ldots, X_{n+1}$ and the maps f_i. We shall prove Proposition 2.90 by strong induction on n: [75]

[75] The induction principle that we are applying here is Theorem 2.60 with $g = 1$ (since $\mathbb{Z}_{\geq 1}$ is the set of all positive integers).

Induction step: Let $m \in \mathbb{Z}_{\geq 1}$. Assume that Proposition 2.90 holds under the condition that $n < m$. We must prove that Proposition 2.90 holds under the condition that $n = m$. In other words, we must prove the following claim:

> *Claim 1:* Let $X_1, X_2, \ldots, X_{m+1}$ be $m+1$ sets. For each $i \in \{1, 2, \ldots, m\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then, every complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$ equals $C(f_m, f_{m-1}, \ldots, f_1)$.

[*Proof of Claim 1:* Let γ be a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$. Thus, we must prove that $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$.

We have $m \in \mathbb{Z}_{\geq 1}$, thus $m \geq 1$. Hence, either $m = 1$ or $m > 1$. Thus, we are in one of the following two cases:

Case 1: We have $m = 1$.

Case 2: We have $m > 1$.

Let us first consider Case 1. In this case, we have $m = 1$. Thus, we have $C(f_m, f_{m-1}, \ldots, f_1) = f_1$ (by the definition of $C(f_m, f_{m-1}, \ldots, f_1)$).

Recall that $m = 1$. Thus, the definition of a "complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$" shows that there is only one complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$, and this is simply the map $f_1 : X_1 \to X_2$. Hence, γ is simply the map $f_1 : X_1 \to X_2$ (since γ is a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$). Thus, $\gamma = f_1 = C(f_m, f_{m-1}, \ldots, f_1)$ (since $C(f_m, f_{m-1}, \ldots, f_1) = f_1$). Thus, $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$ is proven in Case 1.

Now, let us consider Case 2. In this case, we have $m > 1$. Hence, the definition of a "complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$" shows that any complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$ is a map of the form $\alpha \circ \beta$, where

- k is some element of $\{1, 2, \ldots, m-1\}$;
- α is a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_{k+1}$;
- β is a complete parenthesization of $f_k \circ f_{k-1} \circ \cdots \circ f_1$.

Thus, γ is a map of this form (since γ is a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$). In other words, we can write γ in the form $\gamma = \alpha \circ \beta$, where k is some element of $\{1, 2, \ldots, m-1\}$, where α is a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_{k+1}$, and where β is a complete parenthesization of $f_k \circ f_{k-1} \circ \cdots \circ f_1$. Consider these k, α and β.

We have $k \in \{1, 2, \ldots, m-1\}$, thus $k \leq m - 1 < m$. Hence, we can apply Proposition 2.90 to $n = k$ (since we assumed that Proposition 2.90 holds under the condition that $n < m$). We thus conclude that every complete parenthesization of $f_k \circ f_{k-1} \circ \cdots \circ f_1$ equals $C(f_k, f_{k-1}, \ldots, f_1)$. Hence, β equals $C(f_k, f_{k-1}, \ldots, f_1)$ (since β is a complete parenthesization of $f_k \circ f_{k-1} \circ \cdots \circ f_1$). In other words,

$$\beta = C(f_k, f_{k-1}, \ldots, f_1). \tag{136}$$

We have $k \in \{1, 2, \ldots, m-1\}$, thus $k \geq 1$ and therefore $m - \underbrace{k}_{\geq 1} \leq m - 1 < m$. Hence, we can apply Proposition 2.90 to $m-k$, X_{k+i} and f_{k+i} instead of n, X_i and f_i (since we assumed that Proposition 2.90 holds under the condition that $n < m$). We thus conclude that every complete parenthesization of $f_{k+(m-k)} \circ f_{k+(m-k-1)} \circ \cdots \circ f_{k+1}$ equals $C\left(f_{k+(m-k)}, f_{k+(m-k-1)}, \ldots, f_{k+1}\right)$.

Since $k + (m-k) = m$ and $k + (m-k-1) = m-1$, this rewrites as follows: Every complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_{k+1}$ equals $C(f_m, f_{m-1}, \ldots, f_{k+1})$. Thus, α equals $C(f_m, f_{m-1}, \ldots, f_{k+1})$ (since α is a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_{k+1}$). In other words,

$$\alpha = C(f_m, f_{m-1}, \ldots, f_{k+1}). \tag{137}$$

But Proposition 2.89 (applied to $n = m$) yields

$$C(f_m, f_{m-1}, \ldots, f_1) = \underbrace{C(f_m, f_{m-1}, \ldots, f_{k+1})}_{\substack{=\alpha \\ \text{(by (137))}}} \circ \underbrace{C(f_k, f_{k-1}, \ldots, f_1)}_{\substack{=\beta \\ \text{(by (136))}}} = \alpha \circ \beta = \gamma$$

(since $\gamma = \alpha \circ \beta$), so that $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$. Hence, $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$ is proven in Case 2.

We now have shown that $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$ in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, this yields that $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$ always holds.

Now, forget that we fixed γ. We thus have shown that $\gamma = C(f_m, f_{m-1}, \ldots, f_1)$ whenever γ is a complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$. In other words, every complete parenthesization of $f_m \circ f_{m-1} \circ \cdots \circ f_1$ equals $C(f_m, f_{m-1}, \ldots, f_1)$. This proves Claim 1.]

Now, we have proven Claim 1. In other words, we have proven that Proposition 2.90 holds under the condition that $n = m$. This completes the induction step. Hence, Proposition 2.90 is proven by strong induction. □

Proof of Theorem 2.86. Proposition 2.90 shows that every complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ equals $C(f_n, f_{n-1}, \ldots, f_1)$. Thus, all complete parenthesizations of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ are the same map. This proves Theorem 2.86. □

2.13.7. Compositions of multiple maps without parentheses

Definition 2.91. Let n be a positive integer. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. Then, the map $C(f_n, f_{n-1}, \ldots, f_1) : X_1 \to X_{n+1}$ is denoted by $f_n \circ f_{n-1} \circ \cdots \circ f_1$ and called the *composition* of $f_n, f_{n-1}, \ldots, f_1$. This notation $f_n \circ f_{n-1} \circ \cdots \circ f_1$ may conflict with existing notations in two cases:

- In the case when $n = 1$, this notation $f_n \circ f_{n-1} \circ \cdots \circ f_1$ simply becomes f_1, which looks exactly like the map f_1 itself. Fortunately, this conflict of notation is harmless, because the new meaning that we are giving to f_1 in this case (namely, $C(f_n, f_{n-1}, \ldots, f_1) = C(f_1)$) agrees with the map f_1 (because of (130)).
- In the case when $n = 2$, this notation $f_n \circ f_{n-1} \circ \cdots \circ f_1$ simply becomes $f_2 \circ f_1$, which looks exactly like the composition $f_2 \circ f_1$ of the two maps f_2 and f_1. Fortunately, this conflict of notation is harmless, because the new meaning that we are giving to $f_2 \circ f_1$ in this case (namely, $C(f_n, f_{n-1}, \ldots, f_1) = C(f_2, f_1)$) agrees with the latter composition (because of (131)).

Thus, in both cases, the conflict with existing notations is harmless (the conflicting notations actually stand for the same thing).

Remark 2.92. Let n, $X_1, X_2, \ldots, X_{n+1}$ and f_i be as in Definition 2.91. Then, Proposition 2.90 shows that every complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ equals $C(f_n, f_{n-1}, \ldots, f_1)$. In other words, every complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$ equals $f_n \circ f_{n-1} \circ \cdots \circ f_1$ (because $f_n \circ f_{n-1} \circ \cdots \circ f_1$ was defined to be $C(f_n, f_{n-1}, \ldots, f_1)$ in Definition 2.91). In other words, **we can drop the parentheses** in every complete parenthesization of $f_n \circ f_{n-1} \circ \cdots \circ f_1$. For example, for $n = 7$, we get

$$(f_7 \circ (f_6 \circ f_5)) \circ (f_4 \circ ((f_3 \circ f_2) \circ f_1)) = f_7 \circ f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1,$$

and a similar equality for any other complete parenthesization.

Definition 2.91 and Remark 2.92 finally give us the justification to write compositions of multiple maps (like $f_n \circ f_{n-1} \circ \cdots \circ f_1$ for $n \geq 1$) without the need for parentheses. We shall now go one little step further and extend this notation to the case of $n = 0$ – that is, we shall define the composition of **no** maps:

Definition 2.93. Let $n \in \mathbb{N}$. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map. In Definition 2.91, we have defined the composition $f_n \circ f_{n-1} \circ \cdots \circ f_1$ of $f_n, f_{n-1}, \ldots, f_1$ when n is a positive integer. We shall now extend this definition to the case when $n = 0$ (so that it will be defined for all $n \in \mathbb{N}$, not just for all positive integers n). Namely, we extend it by setting

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 = \mathrm{id}_{X_1} \qquad \text{when } n = 0. \tag{138}$$

That is, we say that the composition of 0 maps is the identity map $\mathrm{id}_{X_1} : X_1 \to X_1$. This composition of 0 maps is also known as the *empty composition of maps*. Thus, the empty composition of maps is defined to be id_{X_1}. (This is similar to the well-known conventions that a sum of 0 numbers is 0, and that a product of 0 numbers is 1.)

This definition is slightly dangerous, because it entails that the composition of 0 maps depends on the set X_1, but of course the 0 maps being composed know nothing about this set X_1. Thus, when we speak of an empty composition, we should always specify the set X_1 or ensure that it is clear from the context. (See Definition 2.94 below for an example where it is clear from the context.)

2.13.8. Composition powers

Having defined the composition of n maps, we get the notion of composition powers of maps for free:

Definition 2.94. Let $n \in \mathbb{N}$. Let X be a set. Let $f : X \to X$ be a map. Then, $f^{\circ n}$ shall denote the map

$$\underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times } f} : X \to X.$$

Thus, in particular,

$$f^{\circ 0} = \underbrace{f \circ f \circ \cdots \circ f}_{0 \text{ times } f} = \text{id}_X \tag{139}$$

(by Definition 2.93). Also, $f^{\circ 1} = f$, $f^{\circ 2} = f \circ f$, $f^{\circ 3} = f \circ f \circ f$, etc.

The map $f^{\circ n}$ is called the *n-th composition power* of f (or simply the *n-th power* of f).

Before we study composition powers in detail, let us show a general rule that allows us to "split" compositions of maps:

Theorem 2.95. Let $n \in \mathbb{N}$. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be a map.

(a) We have

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 = (f_n \circ f_{n-1} \circ \cdots \circ f_{k+1}) \circ (f_k \circ f_{k-1} \circ \cdots \circ f_1)$$

for each $k \in \{0, 1, \ldots, n\}$.

(b) If $n \geq 1$, then

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 = f_n \circ (f_{n-1} \circ f_{n-2} \circ \cdots \circ f_1).$$

(c) If $n \geq 1$, then

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 = (f_n \circ f_{n-1} \circ \cdots \circ f_2) \circ f_1.$$

Proof of Theorem 2.95. **(a)** Let $k \in \{0, 1, \ldots, n\}$. We are in one of the following three cases:

Case 1: We have $k = 0$.

Case 2: We have $k = n$.

Case 3: We have neither $k = 0$ nor $k = n$.

(Of course, Cases 1 and 2 overlap when $n = 0$.)

Let us first consider Case 1. In this case, we have $k = 0$. Thus, $f_k \circ f_{k-1} \circ \cdots \circ f_1$ is an empty composition of maps, and therefore equals id_{X_1}. In other words, $f_k \circ f_{k-1} \circ \cdots \circ f_1 = \mathrm{id}_{X_1}$.

On the other hand, $k = 0$, so that $k + 1 = 1$. Hence, $f_n \circ f_{n-1} \circ \cdots \circ f_{k+1} = f_n \circ f_{n-1} \circ \cdots \circ f_1$. Thus,

$$\underbrace{(f_n \circ f_{n-1} \circ \cdots \circ f_{k+1})}_{= f_n \circ f_{n-1} \circ \cdots \circ f_1} \circ \underbrace{(f_k \circ f_{k-1} \circ \cdots \circ f_1)}_{= \mathrm{id}_{X_1}}$$
$$= (f_n \circ f_{n-1} \circ \cdots \circ f_1) \circ \mathrm{id}_{X_1} = f_n \circ f_{n-1} \circ \cdots \circ f_1.$$

In other words,

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 = (f_n \circ f_{n-1} \circ \cdots \circ f_{k+1}) \circ (f_k \circ f_{k-1} \circ \cdots \circ f_1).$$

Hence, Theorem 2.95 **(a)** is proven in Case 1.

Let us next consider Case 2. In this case, we have $k = n$. Thus, $f_n \circ f_{n-1} \circ \cdots \circ f_{k+1}$ is an empty composition of maps, and therefore equals $\mathrm{id}_{X_{n+1}}$. In other words, $f_n \circ f_{n-1} \circ \cdots \circ f_{k+1} = \mathrm{id}_{X_{n+1}}$. Thus,

$$\underbrace{(f_n \circ f_{n-1} \circ \cdots \circ f_{k+1})}_{= \mathrm{id}_{X_{n+1}}} \circ \underbrace{(f_k \circ f_{k-1} \circ \cdots \circ f_1)}_{\substack{= f_n \circ f_{n-1} \circ \cdots \circ f_1 \\ (\text{since } k = n)}}$$
$$= \mathrm{id}_{X_{n+1}} \circ (f_n \circ f_{n-1} \circ \cdots \circ f_1) = f_n \circ f_{n-1} \circ \cdots \circ f_1.$$

In other words,

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 = (f_n \circ f_{n-1} \circ \cdots \circ f_{k+1}) \circ (f_k \circ f_{k-1} \circ \cdots \circ f_1).$$

Hence, Theorem 2.95 **(a)** is proven in Case 2.

Let us finally consider Case 3. In this case, we have neither $k = 0$ nor $k = n$. In other words, we have $k \neq 0$ and $k \neq n$. Combining $k \in \{0, 1, \ldots, n\}$ with $k \neq 0$, we find $k \in \{0, 1, \ldots, n\} \setminus \{0\} \subseteq \{1, 2, \ldots, n\}$. Combining this with $k \neq n$, we find $k \in \{1, 2, \ldots, n\} \setminus \{n\} \subseteq \{1, 2, \ldots, n-1\}$. Hence, $1 \leq k \leq n - 1$, so that $n - 1 \geq 1$ and thus $n \geq 2 \geq 1$. Hence, n is a positive integer. Thus, Proposition 2.89 yields

$$C(f_n, f_{n-1}, \ldots, f_1) = C(f_n, f_{n-1}, \ldots, f_{k+1}) \circ C(f_k, f_{k-1}, \ldots, f_1). \tag{140}$$

Now, k is a positive integer (since $k \in \{1, 2, \ldots, n-1\}$). Hence, Definition 2.91 (applied to k instead of n) yields

$$f_k \circ f_{k-1} \circ \cdots \circ f_1 = C(f_k, f_{k-1}, \ldots, f_1). \tag{141}$$

Also, $n-k$ is an integer satisfying $n - \underbrace{k}_{\leq n-1} \geq n-(n-1) = 1$. Hence, $n-k$ is a positive integer. Thus, Definition 2.91 (applied to $n-k$, X_{k+i} and f_{k+i} instead of n, X_i and f_i) yields

$$f_{k+(n-k)} \circ f_{k+(n-k-1)} \circ \cdots \circ f_{k+1} = C\left(f_{k+(n-k)}, f_{k+(n-k-1)}, \ldots, f_{k+1}\right).$$

In view of $k+(n-k) = n$ and $k+(n-k-1) = n-1$, this rewrites as follows:

$$f_n \circ f_{n-1} \circ \cdots \circ f_{k+1} = C\left(f_n, f_{n-1}, \ldots, f_{k+1}\right). \qquad (142)$$

But n is a positive integer. Thus, Definition 2.91 yields

$$\begin{aligned} f_n \circ f_{n-1} \circ \cdots \circ f_1 &= C\left(f_n, f_{n-1}, \ldots, f_1\right) \\ &= \underbrace{C\left(f_n, f_{n-1}, \ldots, f_{k+1}\right)}_{\substack{= f_n \circ f_{n-1} \circ \cdots \circ f_{k+1} \\ \text{(by (142))}}} \circ \underbrace{C\left(f_k, f_{k-1}, \ldots, f_1\right)}_{\substack{= f_k \circ f_{k-1} \circ \cdots \circ f_1 \\ \text{(by (141))}}} \qquad \text{(by (140))} \\ &= \left(f_n \circ f_{n-1} \circ \cdots \circ f_{k+1}\right) \circ \left(f_k \circ f_{k-1} \circ \cdots \circ f_1\right). \end{aligned}$$

Hence, Theorem 2.95 **(a)** is proven in Case 3.

We have now proven Theorem 2.95 **(a)** in each of the three Cases 1, 2 and 3. Since these three Cases cover all possibilities, we thus conclude that Theorem 2.95 **(a)** always holds.

(b) Assume that $n \geq 1$. Hence, $n-1 \geq 0$, so that $n-1 \in \{0, 1, \ldots, n\}$ (since $n-1 \leq n$). Hence, Theorem 2.95 **(a)** (applied to $k = n-1$) yields

$$\begin{aligned} f_n \circ f_{n-1} \circ \cdots \circ f_1 &= \underbrace{\left(f_n \circ f_{n-1} \circ \cdots \circ f_{(n-1)+1}\right)}_{= f_n \circ f_{n-1} \circ \cdots \circ f_n = f_n} \circ \underbrace{\left(f_{n-1} \circ f_{(n-1)-1} \circ \cdots \circ f_1\right)}_{= f_{n-1} \circ f_{n-2} \circ \cdots \circ f_1} \\ &= f_n \circ \left(f_{n-1} \circ f_{n-2} \circ \cdots \circ f_1\right). \end{aligned}$$

This proves Theorem 2.95 **(b)**.

(c) Assume that $n \geq 1$. Hence, $1 \in \{0, 1, \ldots, n\}$. Hence, Theorem 2.95 **(a)** (applied to $k = 1$) yields

$$\begin{aligned} f_n \circ f_{n-1} \circ \cdots \circ f_1 &= \underbrace{\left(f_n \circ f_{n-1} \circ \cdots \circ f_{1+1}\right)}_{= f_n \circ f_{n-1} \circ \cdots \circ f_2} \circ \underbrace{\left(f_1 \circ f_{1-1} \circ \cdots \circ f_1\right)}_{= f_1} \\ &= \left(f_n \circ f_{n-1} \circ \cdots \circ f_2\right) \circ f_1. \end{aligned}$$

This proves Theorem 2.95 **(c)**. □

We can draw some consequences about composition powers of maps from Theorem 2.95:

Proposition 2.96. Let X be a set. Let $f : X \to X$ be a map. Let n be a positive integer.

(a) We have $f^{\circ n} = f \circ f^{\circ (n-1)}$.

(b) We have $f^{\circ n} = f^{\circ (n-1)} \circ f$.

Proof of Proposition 2.96. The definition of $f^{\circ n}$ yields

$$f^{\circ n} = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times } f}. \qquad (143)$$

The definition of $f^{\circ (n-1)}$ yields

$$f^{\circ (n-1)} = \underbrace{f \circ f \circ \cdots \circ f}_{n-1 \text{ times } f}. \qquad (144)$$

(a) Theorem 2.95 **(b)** (applied to $X_i = X$ and $f_i = f$) yields

$$\underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times } f} = f \circ \left(\underbrace{f \circ f \circ \cdots \circ f}_{n-1 \text{ times } f} \right).$$

In view of (143) and (144), this rewrites as $f^{\circ n} = f \circ f^{\circ (n-1)}$. This proves Proposition 2.96 **(a)**.

(b) Theorem 2.95 **(c)** (applied to $X_i = X$ and $f_i = f$) yields

$$\underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times } f} = \left(\underbrace{f \circ f \circ \cdots \circ f}_{n-1 \text{ times } f} \right) \circ f.$$

In view of (143) and (144), this rewrites as $f^{\circ n} = f^{\circ (n-1)} \circ f$. This proves Proposition 2.96 **(b)**. ☐

Proposition 2.97. Let X be a set. Let $f : X \to X$ be a map.

(a) We have $f^{\circ (a+b)} = f^{\circ a} \circ f^{\circ b}$ for every $a \in \mathbb{N}$ and $b \in \mathbb{N}$.

(b) We have $f^{\circ (ab)} = \left(f^{\circ a}\right)^{\circ b}$ for every $a \in \mathbb{N}$ and $b \in \mathbb{N}$.

Proof of Proposition 2.97. **(a)** Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$. Thus, $a \geq 0$ and $b \geq 0$, so that $0 \leq b \leq a + b$ (since $\underbrace{a}_{\geq 0} + b \geq b$). Hence, $b \in \{0, 1, \ldots, a + b\}$. Thus, Theorem 2.95 **(a)** (applied to $n = a + b$, $X_i = X$, $f_i = f$ and $k = b$) yields

$$\underbrace{f \circ f \circ \cdots \circ f}_{a+b \text{ times } f} = \left(\underbrace{f \circ f \circ \cdots \circ f}_{(a+b)-b \text{ times } f} \right) \circ \left(\underbrace{f \circ f \circ \cdots \circ f}_{b \text{ times } f} \right).$$

In view of $(a+b)-b=a$, this rewrites as

$$\underbrace{f \circ f \circ \cdots \circ f}_{a+b \text{ times } f} = \left(\underbrace{f \circ f \circ \cdots \circ f}_{a \text{ times } f} \right) \circ \left(\underbrace{f \circ f \circ \cdots \circ f}_{b \text{ times } f} \right). \tag{145}$$

But the definition of $f^{\circ a}$ yields

$$f^{\circ a} = \underbrace{f \circ f \circ \cdots \circ f}_{a \text{ times } f}. \tag{146}$$

Also, the definition of $f^{\circ b}$ yields

$$f^{\circ b} = \underbrace{f \circ f \circ \cdots \circ f}_{b \text{ times } f}. \tag{147}$$

Finally, the definition of $f^{\circ(a+b)}$ yields

$$f^{\circ(a+b)} = \underbrace{f \circ f \circ \cdots \circ f}_{a+b \text{ times } f}. \tag{148}$$

In view of these three equalities (146), (147) and (148), we can rewrite the equality (145) as $f^{\circ(a+b)} = f^{\circ a} \circ f^{\circ b}$. This proves Proposition 2.97 **(a)**.

(Alternatively, it is easy to prove Proposition 2.97 **(a)** by induction on a.)

(b) Let $a \in \mathbb{N}$. We claim that

$$f^{\circ(ab)} = \left(f^{\circ a}\right)^{\circ b} \qquad \text{for every } b \in \mathbb{N}. \tag{149}$$

We shall prove (149) by induction on b:

Induction base: We have $a \cdot 0 = 0$ and thus $f^{\circ(a \cdot 0)} = f^{\circ 0} = \text{id}_X$ (by (139)). Comparing this with $\left(f^{\circ a}\right)^{\circ 0} = \text{id}_X$ (which follows from (139), applied to $f^{\circ a}$ instead of f), we obtain $f^{\circ(a \cdot 0)} = \left(f^{\circ a}\right)^{\circ 0}$. In other words, (149) holds for $b = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (149) holds for $b = m$. We must prove that (149) holds for $b = m+1$.

We have assumed that (149) holds for $b = m$. In other words, we have $f^{\circ(am)} = \left(f^{\circ a}\right)^{\circ m}$.

But $m+1$ is a positive integer (since $m+1 > m \geq 0$). Hence, Proposition 2.96 **(b)** (applied to $m+1$ and $f^{\circ a}$ instead of n and f) yields

$$\left(f^{\circ a}\right)^{\circ(m+1)} = \left(f^{\circ a}\right)^{\circ((m+1)-1)} \circ f^{\circ a} = \left(f^{\circ a}\right)^{\circ m} \circ f^{\circ a} \tag{150}$$

(since $(m+1)-1 = m$).

But $a(m+1) = am + a$. Thus,

$$f^{\circ(a(m+1))} = f^{\circ(am+a)} = \underbrace{f^{\circ(am)}}_{=(f^{\circ a})^{\circ m}} \circ f^{\circ a}$$
$$\left(\begin{array}{c}\text{by Proposition 2.97 **(a)**} \\ \text{(applied to } am \text{ and } a \text{ instead of } a \text{ and } b\text{)}\end{array}\right)$$
$$= (f^{\circ a})^{\circ m} \circ f^{\circ a} = (f^{\circ a})^{\circ(m+1)} \qquad \text{(by (150))}.$$

In other words, (149) holds for $b = m+1$. This completes the induction step. Thus, (149) is proven by induction. Hence, Proposition 2.97 **(b)** is proven. □

Note that Proposition 2.97 is similar to the rules of exponents

$$n^{a+b} = n^a n^b \qquad \text{and} \qquad n^{ab} = (n^a)^b$$

that hold for $n \in \mathbb{Q}$ and $a, b \in \mathbb{N}$ (and for various other situations). Can we find similar analogues for other rules of exponents, such as $(mn)^a = m^a n^a$? The simplest analogue one could think of for this rule would be $(f \circ g)^{\circ a} = f^{\circ a} \circ g^{\circ a}$; but this does not hold in general (unless $a \leq 1$). However, it turns out that this does hold if we assume that $f \circ g = g \circ f$ (which is not automatically true, unlike the analogous equality $mn = nm$ for integers). Let us prove this:

Proposition 2.98. Let X be a set. Let $f : X \to X$ and $g : X \to X$ be two maps such that $f \circ g = g \circ f$. Then:
(a) We have $f \circ g^{\circ b} = g^{\circ b} \circ f$ for each $b \in \mathbb{N}$.
(b) We have $f^{\circ a} \circ g^{\circ b} = g^{\circ b} \circ f^{\circ a}$ for each $a \in \mathbb{N}$ and $b \in \mathbb{N}$.
(c) We have $(f \circ g)^{\circ a} = f^{\circ a} \circ g^{\circ a}$ for each $a \in \mathbb{N}$.

Example 2.99. Let us see why the requirement $f \circ g = g \circ f$ is needed in Proposition 2.98:

Let X be the set $\mathbb{Z}$. Let $f : X \to X$ be the map that sends every integer x to $-x$. Let $g : X \to X$ be the map that sends every integer x to $1 - x$. Then, $f^{\circ 2} = \mathrm{id}_X$ (since $f^{\circ 2}(x) = f(f(x)) = -(-x) = x$ for each $x \in X$) and $g^{\circ 2} = \mathrm{id}_X$ (since $g^{\circ 2}(x) = g(g(x)) = 1 - (1 - x) = x$ for each $x \in X$). But the map $f \circ g$ satisfies $(f \circ g)(x) = f(g(x)) = -(1 - x) = x - 1$ for each $x \in X$. Hence, $(f \circ g)^{\circ 2}(x) = (f \circ g)((f \circ g)(x)) = (x - 1) - 1 = x - 2$ for each $x \in X$. Thus, $(f \circ g)^{\circ 2} \neq \mathrm{id}_X$. Comparing this with $\underbrace{f^{\circ 2}}_{=\mathrm{id}_X} \circ \underbrace{g^{\circ 2}}_{=\mathrm{id}_X} = \mathrm{id}_X \circ \mathrm{id}_X = \mathrm{id}_X$, we obtain $(f \circ g)^{\circ 2} \neq f^{\circ 2} \circ g^{\circ 2}$. This shows that Proposition 2.98 **(c)** would not hold without the requirement $f \circ g = g \circ f$.

Proof of Proposition 2.98. **(a)** We claim that

$$f \circ g^{\circ b} = g^{\circ b} \circ f \qquad \text{for each } b \in \mathbb{N}. \tag{151}$$

Indeed, let us prove (151) by induction on b:

Induction base: We have $g^{\circ 0} = \mathrm{id}_X$ (by (139), applied to g instead of f). Hence, $f \circ \underbrace{g^{\circ 0}}_{=\mathrm{id}_X} = f \circ \mathrm{id}_X = f$ and $\underbrace{g^{\circ 0}}_{=\mathrm{id}_X} \circ f = \mathrm{id}_X \circ f = f$. Comparing these two equalities, we obtain $f \circ g^{\circ 0} = g^{\circ 0} \circ f$. In other words, (151) holds for $b = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (151) holds for $b = m$. We must prove that (151) holds for $b = m + 1$.

We have assumed that (151) holds for $b = m$. In other words,

$$f \circ g^{\circ m} = g^{\circ m} \circ f. \tag{152}$$

Proposition 2.82 (applied to $Y = X$, $Z = X$, $W = X$, $c = g$, $b = g^{\circ m}$ and $a = f$) yields

$$(f \circ g^{\circ m}) \circ g = f \circ (g^{\circ m} \circ g). \tag{153}$$

Proposition 2.97 **(a)** (applied to g, m and 1 instead of f, a and b) yields

$$g^{\circ(m+1)} = g^{\circ m} \circ \underbrace{g^{\circ 1}}_{=g} = g^{\circ m} \circ g. \tag{154}$$

Hence,

$$\begin{aligned} f \circ \underbrace{g^{\circ(m+1)}}_{=g^{\circ m} \circ g} &= f \circ (g^{\circ m} \circ g) = \underbrace{(f \circ g^{\circ m})}_{\substack{=g^{\circ m} \circ f \\ \text{(by (152))}}} \circ g \qquad \text{(by (153))} \\ &= (g^{\circ m} \circ f) \circ g = g^{\circ m} \circ \underbrace{(f \circ g)}_{=g \circ f} \\ &\qquad \left(\begin{array}{c} \text{by Proposition 2.82 (applied} \\ \text{to } Y = X,\ Z = X,\ W = X,\ c = g,\ b = f \text{ and } a = g^{\circ m}) \end{array} \right) \\ &= g^{\circ m} \circ (g \circ f). \end{aligned} \tag{155}$$

On the other hand, Proposition 2.82 (applied to $Y = X$, $Z = X$, $W = X$, $c = f$, $b = g$ and $a = g^{\circ m}$) yields

$$(g^{\circ m} \circ g) \circ f = g^{\circ m} \circ (g \circ f).$$

Comparing this with (155), we obtain

$$f \circ g^{\circ(m+1)} = \underbrace{(g^{\circ m} \circ g)}_{\substack{=g^{\circ(m+1)} \\ \text{(by (154))}}} \circ f = g^{\circ(m+1)} \circ f.$$

In other words, (151) holds for $b = m + 1$. This completes the induction step. Thus, (151) is proven by induction.

Therefore, Proposition 2.98 **(a)** follows.

(b) Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$. From $f \circ g = g \circ f$, we obtain $g \circ f = f \circ g$. Hence, Proposition 2.98 **(a)** (applied to g, f and a instead of f, g and b) yields $g \circ f^{\circ a} = f^{\circ a} \circ g$. In other words, $f^{\circ a} \circ g = g \circ f^{\circ a}$. Hence, Proposition 2.98 **(a)** (applied to $f^{\circ a}$ instead of f) yields $f^{\circ a} \circ g^{\circ b} = g^{\circ b} \circ f^{\circ a}$. This proves Proposition 2.98 **(b)**.

(c) We claim that

$$(f \circ g)^{\circ a} = f^{\circ a} \circ g^{\circ a} \qquad \text{for each } a \in \mathbb{N}. \tag{156}$$

Indeed, let us prove (156) by induction on a:

Induction base: From (139), we obtain $f^{\circ 0} = \mathrm{id}_X$ and $g^{\circ 0} = \mathrm{id}_X$ and $(f \circ g)^{\circ 0} = \mathrm{id}_X$. Thus,

$$(f \circ g)^{\circ 0} = \mathrm{id}_X = \underbrace{\mathrm{id}_X}_{= f^{\circ 0}} \circ \underbrace{\mathrm{id}_X}_{= g^{\circ 0}} = f^{\circ 0} \circ g^{\circ 0}.$$

In other words, (156) holds for $a = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (156) holds for $a = m$. We must prove that (156) holds for $a = m + 1$.

We have assumed that (156) holds for $a = m$. In other words,

$$(f \circ g)^{\circ m} = f^{\circ m} \circ g^{\circ m}. \tag{157}$$

But Proposition 2.97 **(a)** (applied to g, m and 1 instead of f, a and b) yields

$$g^{\circ (m+1)} = g^{\circ m} \circ \underbrace{g^{\circ 1}}_{= g} = g^{\circ m} \circ g.$$

The same argument (applied to f instead of g) yields $f^{\circ (m+1)} = f^{\circ m} \circ f$. Hence,

$$\underbrace{f^{\circ (m+1)}}_{= f^{\circ m} \circ f} \circ g^{\circ (m+1)} = (f^{\circ m} \circ f) \circ g^{\circ (m+1)} = f^{\circ m} \circ \left(f \circ g^{\circ (m+1)} \right) \tag{158}$$

(by Proposition 2.82 (applied to $Y = X$, $Z = X$, $W = X$, $c = g^{\circ (m+1)}$, $b = f$ and $a = f^{\circ m}$)).

But Proposition 2.98 **(a)** (applied to $b = m + 1$) yields

$$f \circ g^{\circ (m+1)} = \underbrace{g^{\circ (m+1)}}_{= g^{\circ m} \circ g} \circ f = (g^{\circ m} \circ g) \circ f = g^{\circ m} \circ (g \circ f)$$

(by Proposition 2.82 (applied to $Y = X$, $Z = X$, $W = X$, $c = f$, $b = g$ and $a = g^{\circ m}$)). Hence,

$$f \circ g^{\circ (m+1)} = g^{\circ m} \circ \underbrace{(g \circ f)}_{\substack{= f \circ g \\ (\text{since } f \circ g = g \circ f)}} = g^{\circ m} \circ (f \circ g). \tag{159}$$

On the other hand, Proposition 2.97 **(a)** (applied to $f \circ g$, m and 1 instead of f, a and b) yields

$$(f \circ g)^{\circ(m+1)} = \underbrace{(f \circ g)^{\circ m}}_{\substack{= f^{\circ m} \circ g^{\circ m} \\ \text{(by (157))}}} \circ \underbrace{(f \circ g)^{\circ 1}}_{= f \circ g} = (f^{\circ m} \circ g^{\circ m}) \circ (f \circ g) = f^{\circ m} \circ (g^{\circ m} \circ (f \circ g))$$

(by Proposition 2.82 (applied to $Y = X$, $Z = X$, $W = X$, $c = f \circ g$, $b = g^{\circ m}$ and $a = f^{\circ m}$)). Hence,

$$(f \circ g)^{\circ(m+1)} = f^{\circ m} \circ \underbrace{(g^{\circ m} \circ (f \circ g))}_{\substack{= f \circ g^{\circ(m+1)} \\ \text{(by (159))}}} = f^{\circ m} \circ \left(f \circ g^{\circ(m+1)}\right) = f^{\circ(m+1)} \circ g^{\circ(m+1)}$$

(by (158)). In other words, (156) holds for $a = m + 1$. This completes the induction step. Thus, (156) is proven by induction. Therefore, Proposition 2.98 **(c)** follows. □

Remark 2.100. In our above proof of Proposition 2.98, we have not used the notation $f_n \circ f_{n-1} \circ \cdots \circ f_1$ introduced in Definition 2.91, but instead relied on parentheses and compositions of two maps (i.e., we have never composed more than two maps at the same time). Thus, for example, in the proof of Proposition 2.98 **(a)**, we wrote "$(g^{\circ m} \circ g) \circ f$" and "$g^{\circ m} \circ (g \circ f)$" rather than "$g^{\circ m} \circ g \circ f$". But Remark 2.92 says that we could have just as well dropped all the parentheses. This would have saved us the trouble of explicitly applying Proposition 2.82 (since if we drop all parentheses, then there is no difference between "$(g^{\circ m} \circ g) \circ f$" and "$g^{\circ m} \circ (g \circ f)$" any more). This way, the induction step in the proof of Proposition 2.98 **(a)** could have been made much shorter:

Induction step (second version): Let $m \in \mathbb{N}$. Assume that (151) holds for $b = m$. We must prove that (151) holds for $b = m + 1$.

We have assumed that (151) holds for $b = m$. In other words,

$$f \circ g^{\circ m} = g^{\circ m} \circ f. \tag{160}$$

Proposition 2.97 **(a)** (applied to g, m and 1 instead of f, a and b) yields $g^{\circ(m+1)} = g^{\circ m} \circ \underbrace{g^{\circ 1}}_{= g} = g^{\circ m} \circ g$. Hence,

$$f \circ \underbrace{g^{\circ(m+1)}}_{= g^{\circ m} \circ g} = \underbrace{f \circ g^{\circ m}}_{\substack{= g^{\circ m} \circ f \\ \text{(by (160))}}} \circ g = g^{\circ m} \circ \underbrace{f \circ g}_{= g \circ f} = \underbrace{g^{\circ m} \circ g}_{= g^{\circ(m+1)}} \circ f = g^{\circ(m+1)} \circ f.$$

In other words, (151) holds for $b = m + 1$. This completes the induction step.

Similarly, we can simplify the proof of Proposition 2.98 **(c)** by dropping the parentheses. (The details are left to the reader.)

2.13.9. Composition of invertible maps

The composition of two invertible maps is always invertible, and its inverse can be computed by the following formula:

Proposition 2.101. Let X, Y and Z be three sets. Let $b : X \to Y$ and $a : Y \to Z$ be two invertible maps. Then, the map $a \circ b : X \to Z$ is invertible as well, and its inverse is

$$(a \circ b)^{-1} = b^{-1} \circ a^{-1}.$$

Proof of Proposition 2.101. As we have explained in Definition 2.91, we can drop the parentheses when composing several maps. This will allow us to write expressions like $b^{-1} \circ a^{-1} \circ a \circ b$ without specifying where parentheses should be placed, and then pretending that they are placed wherever we would find them most convenient.

The equalities

$$\left(b^{-1} \circ a^{-1}\right) \circ (a \circ b) = b^{-1} \circ \underbrace{a^{-1} \circ a}_{=\mathrm{id}_Y} \circ b = b^{-1} \circ b = \mathrm{id}_X$$

and

$$(a \circ b) \circ \left(b^{-1} \circ a^{-1}\right) = a \circ \underbrace{b \circ b^{-1}}_{=\mathrm{id}_Y} \circ a^{-1} = a \circ a^{-1} = \mathrm{id}_Z$$

show that the map $b^{-1} \circ a^{-1}$ is an inverse of $a \circ b$. Thus, the map $a \circ b$ has an inverse (namely, $b^{-1} \circ a^{-1}$). In other words, the map $a \circ b$ is invertible. Its inverse is $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$ (since $b^{-1} \circ a^{-1}$ is an inverse of $a \circ b$). This completes the proof of Proposition 2.101. □

By a straightforward induction, we can extend Proposition 2.101 to compositions of n invertible maps:

Proposition 2.102. Let $n \in \mathbb{N}$. Let $X_1, X_2, \ldots, X_{n+1}$ be $n+1$ sets. For each $i \in \{1, 2, \ldots, n\}$, let $f_i : X_i \to X_{i+1}$ be an invertible map. Then, the map $f_n \circ f_{n-1} \circ \cdots \circ f_1 : X_1 \to X_{n+1}$ is invertible as well, and its inverse is

$$(f_n \circ f_{n-1} \circ \cdots \circ f_1)^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_n^{-1}.$$

Exercise 2.2. Prove Proposition 2.102.

In particular, Proposition 2.102 shows that any composition of invertible maps is invertible. Since invertible maps are the same as bijective maps, we can rewrite this as follows: Any composition of bijective maps is bijective.

2.14. General commutativity for addition of numbers

2.14.1. The setup and the problem

Throughout Section 2.14, we let $\mathbb{A}$ be one of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. The elements of $\mathbb{A}$ will be simply called *numbers*.

There is an analogue of Proposition 2.82 for numbers:

Proposition 2.103. Let a, b and c be three numbers (i.e., elements of $\mathbb{A}$). Then, $(a+b)+c = a+(b+c)$.

Proposition 2.103 is known as the *associativity of addition* (in $\mathbb{A}$), and is fundamental; its proof can be found in any textbook on the construction of the number system[76].

In Section 2.13, we have used Proposition 2.82 to show that we can "drop the parentheses" in a composition $f_n \circ f_{n-1} \circ \cdots \circ f_1$ of maps (i.e., all possible complete parenthesizations of this composition are actually the same map). Likewise, we can use Proposition 2.103 to show that we can "drop the parentheses" in a sum $a_1 + a_2 + \cdots + a_n$ of numbers (i.e., all possible complete parenthesizations of this sum are actually the same number). For example, if a, b, c, d are four numbers, then the complete parenthesizations of $a+b+c+d$ are

$$((a+b)+c)+d, \qquad (a+(b+c))+d, \qquad (a+b)+(c+d),$$
$$a+((b+c)+d), \qquad a+(b+(c+d)),$$

and all of these five complete parenthesizations are the same number.

However, numbers behave better than maps. In particular, along with Proposition 2.103, they satisfy another law that maps (generally) don't satisfy:

Proposition 2.104. Let a and b be two numbers (i.e., elements of $\mathbb{A}$). Then, $a+b = b+a$.

Proposition 2.104 is known as the *commutativity of addition* (in $\mathbb{A}$), and again is a fundamental result whose proofs are found in standard textbooks[77].

Furthermore, numbers can **always** be added, whereas maps can only be composed if the domain of one is the codomain of the other. Thus, when we want to take the sum of n numbers $a_1, a_2, \ldots, a_n$, we can not only choose where to put the parentheses, but also in what order the numbers should appear in the sum. It turns

[76]For example, Proposition 2.103 is proven in [Swanso18, Theorem 3.2.3 (3)] for the case when $\mathbb{A} = \mathbb{N}$; in [Swanso18, Theorem 3.5.4 (3)] for the case when $\mathbb{A} = \mathbb{Z}$; in [Swanso18, Theorem 3.6.4 (3)] for the case when $\mathbb{A} = \mathbb{Q}$; in [Swanso18, Theorem 3.7.10] for the case when $\mathbb{A} = \mathbb{R}$; in [Swanso18, Theorem 3.9.3] for the case when $\mathbb{A} = \mathbb{C}$.

[77]For example, Proposition 2.104 is proven in [Swanso18, Theorem 3.2.3 (4)] for the case when $\mathbb{A} = \mathbb{N}$; in [Swanso18, Theorem 3.5.4 (4)] for the case when $\mathbb{A} = \mathbb{Z}$; in [Swanso18, Theorem 3.6.4 (4)] for the case when $\mathbb{A} = \mathbb{Q}$; in [Swanso18, Theorem 3.7.10] for the case when $\mathbb{A} = \mathbb{R}$; in [Swanso18, Theorem 3.9.3] for the case when $\mathbb{A} = \mathbb{C}$.

out that neither of these choices affects the result. For example, if a, b, c are three numbers, then all 12 possible sums

$$\begin{array}{llll} (a+b)+c, & a+(b+c), & (a+c)+b, & a+(c+b), \\ (b+a)+c, & b+(a+c), & (b+c)+a, & b+(c+a), \\ (c+a)+b, & c+(a+b), & (c+b)+a, & c+(b+a) \end{array}$$

are actually the same number. The reader can easily verify this for three numbers a, b, c (using Proposition 2.103 and Proposition 2.104), but of course the general case (with n numbers) is more difficult. The independence of the result on the parenthesization can be proven using the same arguments that we gave in Section 2.13 (except that the $\circ$ symbol is now replaced by $+$), but the independence on the order cannot easily be shown (or even stated) in this way.

Thus, we shall proceed differently: We shall rigorously define the sum of n numbers without specifying an order in which they are added or using parentheses. Unlike the composition of n maps, which was defined for an *ordered list* of n maps, we shall define the sum of n numbers for a *family* of n numbers (see the next subsection for the definition of a "family"). Families don't come with an ordering chosen in advance, so we cannot single out any specific ordering for use in the definition. Thus, the independence on the order will be baked right into the definition.

Different solutions to the problem of formalizing the concept of the sum of n numbers can be found in [Bourba74, Chapter 1, §1.5][78] and in [GalQua18, §3.3].

2.14.2. Families

Let us first define what we mean by a "family" of n numbers. More generally, we can define a family of elements of any set, or even a family of elements of **different** sets. To motivate the definition, we first recall a concept of an n-tuple:

Remark 2.105. Let $n \in \mathbb{N}$.

(a) Let A be a set. Then, to specify an *n-tuple of elements of A* means specifying an element a_i of A for each $i \in \{1, 2, \ldots, n\}$. This n-tuple is then denoted by $(a_1, a_2, \ldots, a_n)$ or by $(a_i)_{i \in \{1,2,\ldots,n\}}$. For each $i \in \{1, 2, \ldots, n\}$, we refer to a_i as the *i-th entry* of this n-tuple.

The set of all n-tuples of elements of A is denoted by A^n or by $A^{\times n}$; it is called the *n-th Cartesian power* of the set A.

(b) More generally, we can define n-tuples of elements from **different** sets: For each $i \in \{1, 2, \ldots, n\}$, let A_i be a set. Then, to specify an *n-tuple of elements of $A_1, A_2, \ldots, A_n$* means specifying an element a_i of A_i for each $i \in \{1, 2, \ldots, n\}$. This n-tuple is (again) denoted by $(a_1, a_2, \ldots, a_n)$ or by $(a_i)_{i \in \{1,2,\ldots,n\}}$. For each $i \in \{1, 2, \ldots, n\}$, we refer to a_i as the *i-th entry* of this n-tuple.

[78] Bourbaki, in [Bourba74, Chapter 1, §1.5], define something more general than a sum of n numbers: They define the "composition" of a finite family of elements of a commutative magma. The sum of n numbers is a particular case of this concept when the magma is the set $\mathbb{A}$ (endowed with its addition).

The set of all n-tuples of elements of $A_1, A_2, \ldots, A_n$ is denoted by $A_1 \times A_2 \times \cdots \times A_n$ or by $\prod_{i=1}^{n} A_i$; it is called the *Cartesian product* of the n sets $A_1, A_2, \ldots, A_n$. These n sets $A_1, A_2, \ldots, A_n$ are called the *factors* of this Cartesian product.

Example 2.106. **(a)** The 3-tuple $(7, 8, 9)$ is a 3-tuple of elements of $\mathbb{N}$, and also a 3-tuple of elements of $\mathbb{Z}$. It can also be written in the form $(6+i)_{i \in \{1,2,3\}}$. Thus, $(6+i)_{i \in \{1,2,3\}} = (6+1, 6+2, 6+3) = (7, 8, 9) \in \mathbb{N}^3$ and also $(6+i)_{i \in \{1,2,3\}} \in \mathbb{Z}^3$.

(b) The 5-tuple $(\{1\}, \{2\}, \{3\}, \varnothing, \mathbb{N})$ is a 5-tuple of elements of the powerset of $\mathbb{N}$ (since $\{1\}, \{2\}, \{3\}, \varnothing, \mathbb{N}$ are subsets of $\mathbb{N}$, thus elements of the powerset of $\mathbb{N}$).

(c) The 0-tuple $()$ can be viewed as a 0-tuple of elements of **any** set A.

(d) If we let $[n]$ be the set $\{1, 2, \ldots, n\}$ for each $n \in \mathbb{N}$, then $(1, 2, 2, 3, 3)$ is a 5-tuple of elements of $[1], [2], [3], [4], [5]$ (because $1 \in [1]$, $2 \in [2]$, $2 \in [3]$, $3 \in [4]$ and $3 \in [5]$). In other words, $(1, 2, 2, 3, 3) \in [1] \times [2] \times [3] \times [4] \times [5]$.

(e) A 2-tuple is the same as an ordered pair. A 3-tuple is the same as an ordered triple. A 1-tuple of elements of a set A is "almost" the same as a single element of A; more precisely, there is a bijection

$$A \to A^1, \qquad a \mapsto (a)$$

from A to the set of 1-tuples of elements of A.

The notation "$(a_i)_{i \in \{1,2,\ldots,n\}}$" in Remark 2.105 should be pronounced as "the n-tuple whose i-th entry is a_i for each $i \in \{1, 2, \ldots, n\}$". The letter "$i$" is used as a variable in this notation (similar to the "i" in the expression "$\sum_{i=1}^{n} i$" or in the expression "the map $\mathbb{N} \to \mathbb{N}$, $i \mapsto i+1$" or in the expression "for all $i \in \mathbb{N}$, we have $i + 1 > i$"); it does not refer to any specific element of $\{1, 2, \ldots, n\}$. As usual, it does not matter which letter we are using for this variable (as long as it does not already have a different meaning); thus, for example, the 3-tuples $(6+i)_{i \in \{1,2,3\}}$ and $(6+j)_{j \in \{1,2,3\}}$ and $(6+x)_{x \in \{1,2,3\}}$ are all identical (and equal $(7, 8, 9)$).

We also note that the "$\prod$" sign in Remark 2.105 **(b)** has a different meaning than the "$\prod$" sign in Section 1.4. The former stands for a Cartesian product of sets, whereas the latter stands for a product of numbers. In particular, a product $\prod_{i=1}^{n} a_i$ of numbers does not change when its factors are swapped, whereas a Cartesian product $\prod_{i=1}^{n} A_i$ of sets does. (In particular, if A and B are two sets, then $A \times B$ and $B \times A$ are different sets in general. The 2-tuple $(1, -1)$ belongs to $\mathbb{N} \times \mathbb{Z}$, but not to $\mathbb{Z} \times \mathbb{N}$.)

Thus, the purpose of an n-tuple is storing several elements (possibly of different sets) in one "container". This is a highly useful notion, but sometimes one wants a more general concept, which can store several elements but not necessarily organized in a "linear order". For example, assume you want to store four integers a, b, c, d in the form of a rectangular table $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ (also known as a "2×2-table of integers"). Such a table doesn't have a well-defined "1-st entry" or "2-nd entry" (unless you agree on a specific order in which you read it); instead, it makes sense to speak of a "$(1,2)$-th entry" (i.e., the entry in row 1 and column 2, which is b) or of a "$(2,2)$-th entry" (i.e., the entry in row 2 and column 2, which is d). Thus, such tables work similarly to n-tuples, but they are "indexed" by pairs (i, j) of appropriate integers rather than by the numbers $1, 2, \ldots, n$.

The concept of a "family" generalizes both n-tuples and rectangular tables: It allows the entries to be indexed by the elements of an arbitrary (possibly infinite) set I instead of the numbers $1, 2, \ldots, n$. Here is its definition (watch the similarities to Remark 2.105):

Definition 2.107. Let I be a set.

(a) Let A be a set. Then, to specify an *I-family of elements of A* means specifying an element a_i of A for each $i \in I$. This I-family is then denoted by $(a_i)_{i \in I}$. For each $i \in I$, we refer to a_i as the *i-th entry* of this I-family. (Unlike the case of n-tuples, there is no notation like $(a_1, a_2, \ldots, a_n)$ for I-families, because there is no natural way in which their entries should be listed.)

An I-family of elements of A is also called an *A-valued I-family*.

The set of all I-families of elements of A is denoted by A^I or by $A^{\times I}$. (Note that the notation A^I is also used for the set of all maps from I to A. But this set is more or less the same as the set of all I-families of elements of A; see Remark 2.109 below for the details.)

(b) More generally, we can define I-families of elements from **different** sets: For each $i \in I$, let A_i be a set. Then, to specify an *I-family of elements of $(A_i)_{i \in I}$* means specifying an element a_i of A_i for each $i \in I$. This I-family is (again) denoted by $(a_i)_{i \in I}$. For each $i \in I$, we refer to a_i as the *i-th entry* of this I-family.

The set of all I-families of elements of $(A_i)_{i \in I}$ is denoted by $\prod_{i \in I} A_i$.

The word "I-family" (without further qualifications) means an I-family of elements of $(A_i)_{i \in I}$ for some sets A_i.

The word "family" (without further qualifications) means an I-family for some set I.

Example 2.108. **(a)** The family $(6 + i)_{i \in \{0,3,5\}}$ is a $\{0, 3, 5\}$-family of elements of $\mathbb{N}$ (that is, an $\mathbb{N}$-valued $\{0, 3, 5\}$-family). It has three entries: Its 0-th entry is $6 + 0 = 6$; its 3-rd entry is $6 + 3 = 9$; its 5-th entry is $6 + 5 = 11$. Of course, this family is also a $\{0, 3, 5\}$-family of elements of $\mathbb{Z}$. If we squint hard enough, we can pretend that this family is simply the 3-tuple $(6, 9, 11)$; but this is not advisable, and also does not extend to situations in which there is no natural order on the set I.

(b) Let X be the set $\{\text{"cat"}, \text{"chicken"}, \text{"dog"}\}$ consisting of three words. Then, we can define an X-family $(a_i)_{i \in X}$ of elements of $\mathbb{N}$ by setting

$$a_{\text{"cat"}} = 4, \qquad a_{\text{"chicken"}} = 2, \qquad a_{\text{"dog"}} = 4.$$

This family has 3 entries, which are 4, 2 and 4; but there is no natural order on the set X, so we cannot identify it with a 3-tuple.

We can also rewrite this family as

$$(\text{the number of legs of a typical specimen of animal } i)_{i \in X}.$$

Of course, not every family will have a description like this; sometimes a family is just a choice of elements without any underlying pattern.

(c) If I is the empty set $\varnothing$, and if A is any set, then there is exactly one I-family of elements of A; namely, the *empty family*. Indeed, specifying such a family means specifying no elements at all, and there is just one way to do that. We can denote the empty family by $()$, just like the empty 0-tuple.

(d) The family $(|i|)_{i \in \mathbb{Z}}$ is a $\mathbb{Z}$-family of elements of $\mathbb{N}$ (because $|i|$ is an element of $\mathbb{N}$ for each $i \in \mathbb{Z}$). It can also be regarded as a $\mathbb{Z}$-family of elements of $\mathbb{Z}$.

(e) If I is the set $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, and if A is any set, then an I-family $(a_i)_{i \in \{1,2,\ldots,n\}}$ of elements of A is the same as an n-tuple of elements of A. The same holds for families and n-tuples of elements from different sets. Thus, any n sets $A_1, A_2, \ldots, A_n$ satisfy $\prod_{i \in \{1,2,\ldots,n\}} A_i = \prod_{i=1}^{n} A_i$.

The notation "$(a_i)_{i \in I}$" in Definition 2.107 should be pronounced as "the I-family whose i-th entry is a_i for each $i \in I$". The letter "i" is used as a variable in this notation (similar to the "i" in the expression "$\sum_{i=1}^{n} i$"); it does not refer to any specific element of I. As usual, it does not matter which letter we are using for this variable (as long as it does not already have a different meaning); thus, for example, the $\mathbb{Z}$-families $(|i|)_{i \in \mathbb{Z}}$ and $(|p|)_{p \in \mathbb{Z}}$ and $(|w|)_{w \in \mathbb{Z}}$ are all identical.

Remark 2.109. Let I and A be two sets. What is the difference between an A-valued I-family and a map from I to A ? Both of these objects consist of a choice of an element of A for each $i \in I$.

The main difference is terminological: e.g., when we speak of a family, the elements of A that constitute it are called its "entries", whereas for a map they are called its "images" or "values". Also, the notations for them are different: The A-valued I-family $(a_i)_{i \in I}$ corresponds to the map $I \to A,\ i \mapsto a_i$.

There is also another, subtler difference: A map from I to A "knows" what the set A is (so that, for example, the maps $\mathbb{N} \to \mathbb{N},\ i \mapsto i$ and $\mathbb{N} \to \mathbb{Z},\ i \mapsto i$ are considered different, even though they map every element of $\mathbb{N}$ to the same value); but an A-valued I-family does not "know" what the set A is (so that, for example, the $\mathbb{N}$-valued $\mathbb{N}$-family $(i)_{i \in \mathbb{N}}$ is considered identical with the $\mathbb{Z}$-valued $\mathbb{N}$-family $(i)_{i \in \mathbb{N}}$). This matters occasionally when one wants to consider maps or families for different sets simultaneously; it is not relevant if we just work with A-valued I-families (or maps from I to A) for two fixed sets I and A. And either way, these conventions are not universal across the mathematical literature; for some authors, maps from I to A do not "know" what A is, whereas other authors want families to "know" this too.

What is certainly true, independently of any conventions, is the following fact: If I and A are two sets, then the map

$$\{\text{maps from } I \text{ to } A\} \to \{A\text{-valued } I\text{-families}\},$$
$$f \mapsto (f(i))_{i \in I}$$

is bijective. (Its inverse map sends every A-valued I-family $(a_i)_{i \in I}$ to the map $I \to A$, $i \mapsto a_i$.) Thus, there is little harm in equating $\{\text{maps from } I \text{ to } A\}$ with $\{A\text{-valued } I\text{-families}\}$.

We already know from Example 2.108 **(e)** that n-tuples are a particular case of families; the same holds for rectangular tables:

Definition 2.110. Let A be a set. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then, an *$n \times m$-table* of elements of A means an A-valued $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$-family. According to Remark 2.109, this is tantamount to saying that an $n \times m$-table of elements of A means a map from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to A, except for notational differences (such as referring to the elements that constitute the $n \times m$-table as "entries" rather than "values") and for the fact that an $n \times m$-table does not "know" A (whereas a map would do).

In future chapters, we shall consider "$n \times m$-matrices", which are defined as maps from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to A rather than as A-valued $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$-families. We shall keep using the same notations for them as for $n \times m$-tables, but unlike $n \times m$-tables, they will "know" A (that is, two $n \times m$-matrices with the same entries but different sets A will be considered different). Anyway, this difference is minor.

2.14.3. A desirable definition

We now know what an $\mathbb{A}$-valued S-family is (for some set S): It is just a way of choosing some element of $\mathbb{A}$ for each $s \in S$. When this element is called a_s, the S-family is called $(a_s)_{s \in S}$.

We now want to define the sum of an $\mathbb{A}$-valued S-family $(a_s)_{s \in S}$ when the set S is finite. Actually, we have already seen a definition of this sum (which is called $\sum\limits_{s \in S} a_s$) in Section 1.4. The only problem with that definition is that we don't know yet that it is legitimate. Let us nevertheless recall it (rewriting it using the notion of an $\mathbb{A}$-valued S-family):

Definition 2.111. If S is a finite set, and if $(a_s)_{s\in S}$ is an $\mathbb{A}$-valued S-family, then we want to define the number $\sum_{s\in S} a_s$. We define this number by recursion on $|S|$ as follows:

- If $|S| = 0$, then $\sum_{s\in S} a_s$ is defined to be 0.
- Let $n \in \mathbb{N}$. Assume that we have defined $\sum_{s\in S} a_s$ for every finite set S with $|S| = n$ and any $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$. Now, if S is a finite set with $|S| = n+1$, and if $(a_s)_{s\in S}$ is any $\mathbb{A}$-valued S-family, then $\sum_{s\in S} a_s$ is defined by picking any $t \in S$ and setting
$$\sum_{s\in S} a_s = a_t + \sum_{s\in S\setminus\{t\}} a_s. \qquad (161)$$

As we already observed in Section 1.4, it is not obvious that this definition is legitimate: The right hand side of (161) is defined using a choice of t, but we want our value of $\sum_{s\in S} a_s$ to depend only on S and $(a_s)_{s\in S}$ (not on some arbitrarily chosen $t \in S$). Thus, we cannot use this definition yet. Our main goal in this section is to prove that it is indeed legitimate.

2.14.4. The set of all possible sums

There are two ways to approach this goal. One is to prove the legitimacy of Definition 2.111 by strong induction on $|S|$; the statement $\mathcal{A}(n)$ that we would be proving for each $n \in \mathbb{N}$ here would be saying that Definition 2.111 is legitimate for all finite sets S satisfying $|S| = n$. This is not hard, but conceptually confusing, as it would require us to use Definition 2.111 for **some** sets S while its legitimacy for other sets S is yet unproven.

We prefer to proceed in a different way: We shall first define a set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ for any $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$; this set shall consist (roughly speaking) of "all possible values that $\sum_{s\in S} a_s$ could have according to Definition 2.111". This set will be defined recursively, more or less following Definition 2.111, but instead of relying on a choice of **some** $t \in S$, it will use **all** possible elements $t \in S$. (See Definition 2.112 for the precise definition.) Unlike $\sum_{s\in S} a_s$ itself, it will be a set of numbers, not a single number; however, it has the advantage that the legitimacy of its definition will be immediately obvious. Then, we will prove (in Theorem 2.114) that this set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is actually a 1-element set; this will allow us to define $\sum_{s\in S} a_s$ to be the unique element of $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ for any $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$ (see Definition 2.116). Then, we will retroactively legitimize Definition 2.111 by showing that Definition 2.111 leads to the same value of $\sum_{s\in S} a_s$ as Definition 2.116 (no matter which $t \in S$ is chosen). Having thus justified Definition 2.111, we will forget about the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ and about Definition 2.116.

In later subsections, we shall prove some basic properties of sums.

Let us define the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$, as promised:

Definition 2.112. If S is a finite set, and if $(a_s)_{s\in S}$ is an $\mathbb{A}$-valued S-family, then we want to define the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ of numbers. We define this set by recursion on $|S|$ as follows:

- If $|S| = 0$, then $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is defined to be $\{0\}$.
- Let $n \in \mathbb{N}$. Assume that we have defined $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ for every finite set S with $|S| = n$ and any $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$. Now, if S is a finite set with $|S| = n+1$, and if $(a_s)_{s\in S}$ is any $\mathbb{A}$-valued S-family, then $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is defined by

$$\begin{aligned}&\operatorname{Sums}\left((a_s)_{s\in S}\right)\\&=\left\{a_t + b \ \mid \ t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)\right\}.\end{aligned} \tag{162}$$

(The sets $\operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)$ on the right hand side of this equation are well-defined, because for each $t \in S$, we have $|S \setminus \{t\}| = |S| - 1 = n$ (since $|S| = n+1$), and therefore $\operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)$ is well-defined by our assumption.)

Example 2.113. Let S be a finite set. Let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Let us see what Definition 2.112 says when S has only few elements:

(a) If $S = \varnothing$, then

$$\operatorname{Sums}\left((a_s)_{s\in\varnothing}\right) = \{0\} \tag{163}$$

(directly by Definition 2.112, since $|S| = |\varnothing| = 0$ in this case).

(b) If $S = \{x\}$ for some element x, then Definition 2.112 yields

$$\begin{aligned}&\operatorname{Sums}\left((a_s)_{s\in\{x\}}\right)\\&=\left\{a_t + b \ \mid \ t \in \{x\} \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in\{x\}\setminus\{t\}}\right)\right\}\\&=\left\{a_x + b \ \mid \ b \in \operatorname{Sums}\left((a_s)_{s\in\{x\}\setminus\{x\}}\right)\right\} \qquad (\text{since the only } t \in \{x\} \text{ is } x)\\&=\left\{a_x + b \ \mid \ b \in \underbrace{\operatorname{Sums}\left((a_s)_{s\in\varnothing}\right)}_{=\{0\}}\right\} \qquad (\text{since } \{x\} \setminus \{x\} = \varnothing)\\&=\{a_x + b \ \mid \ b \in \{0\}\} = \{a_x + 0\} = \{a_x\}.\end{aligned} \tag{164}$$

(c) If $S = \{x, y\}$ for two distinct elements x and y, then Definition 2.112 yields

$$\begin{aligned}
&\operatorname{Sums}\left((a_s)_{s\in\{x,y\}}\right) \\
&= \left\{a_t + b \mid t \in \{x, y\} \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in\{x,y\}\setminus\{t\}}\right)\right\} \\
&= \left\{a_x + b \mid b \in \operatorname{Sums}\left((a_s)_{s\in\{x,y\}\setminus\{x\}}\right)\right\} \\
&\qquad \cup \left\{a_y + b \mid b \in \operatorname{Sums}\left((a_s)_{s\in\{x,y\}\setminus\{y\}}\right)\right\} \\
&= \left\{a_x + b \mid b \in \underbrace{\operatorname{Sums}\left((a_s)_{s\in\{y\}}\right)}_{\substack{=\{a_y\} \\ \text{(by (164), applied to } y \text{ instead of } x)}}\right\} \\
&\qquad \cup \left\{a_y + b \mid b \in \underbrace{\operatorname{Sums}\left((a_s)_{s\in\{x\}}\right)}_{\substack{=\{a_x\} \\ \text{(by (164))}}}\right\} \\
&\qquad \left(\text{since } \{x,y\} \setminus \{x\} = \{y\} \text{ and } \{x,y\} \setminus \{y\} = \{x\}\right) \\
&= \underbrace{\left\{a_x + b \mid b \in \{a_y\}\right\}}_{=\{a_x + a_y\}} \cup \underbrace{\left\{a_y + b \mid b \in \{a_x\}\right\}}_{=\{a_y + a_x\}} \\
&= \{a_x + a_y\} \cup \{a_y + a_x\} = \{a_x + a_y, a_y + a_x\} = \{a_x + a_y\}
\end{aligned}$$

(since $a_y + a_x = a_x + a_y$).

(d) Similar reasoning shows that if $S = \{x, y, z\}$ for three distinct elements x, y and z, then

$$\operatorname{Sums}\left((a_s)_{s\in\{x,y,z\}}\right) = \left\{a_x + (a_y + a_z), a_y + (a_x + a_z), a_z + (a_x + a_y)\right\}.$$

It is not hard to check (using Proposition 2.103 and Proposition 2.104) that the three elements $a_x + (a_y + a_z)$, $a_y + (a_x + a_z)$ and $a_z + (a_x + a_y)$ of this set are equal, so we may call them $a_x + a_y + a_z$; thus, we can rewrite this equality as

$$\operatorname{Sums}\left((a_s)_{s\in\{x,y,z\}}\right) = \{a_x + a_y + a_z\}.$$

(e) Going further, we can see that if $S = \{x, y, z, w\}$ for four distinct elements x, y, z and w, then

$$\begin{aligned}
\operatorname{Sums}\left((a_s)_{s\in\{x,y,z,w\}}\right) = \{&a_x + (a_y + a_z + a_w), a_y + (a_x + a_z + a_w), \\
&a_z + (a_x + a_y + a_w), a_w + (a_x + a_y + a_z)\}.
\end{aligned}$$

Again, it is not hard to prove that

$$a_x + (a_y + a_z + a_w) = a_y + (a_x + a_z + a_w) \\ = a_z + (a_x + a_y + a_w) = a_w + (a_x + a_y + a_z),$$

and thus the set $\operatorname{Sums}\left((a_s)_{s\in\{x,y,z,w\}}\right)$ is again a 1-element set, whose unique element can be called $a_x + a_y + a_z + a_w$.

These examples suggest that the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ should always be a 1-element set. This is precisely what we are going to claim now:

Theorem 2.114. If S is a finite set, and if $(a_s)_{s\in S}$ is an $\mathbb{A}$-valued S-family, then the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is a 1-element set.

2.14.5. The set of all possible sums is a 1-element set: proof

Before we step to the proof of Theorem 2.114, we observe an almost trivial lemma:

Lemma 2.115. Let a, b and c be three numbers (i.e., elements of $\mathbb{A}$). Then, $a + (b + c) = b + (a + c)$.

Proof of Lemma 2.115. Proposition 2.103 (applied to b and a instead of a and b) yields $(b + a) + c = b + (a + c)$. Also, Proposition 2.103 yields $(a + b) + c = a + (b + c)$. Hence,

$$a + (b + c) = \underbrace{(a + b)}_{\substack{=b+a \\ \text{(by Proposition 2.104)}}} + c = (b + a) + c = b + (a + c).$$

This proves Lemma 2.115. □

Proof of Theorem 2.114. We shall prove Theorem 2.114 by strong induction on $|S|$:

Let $m \in \mathbb{N}$. Assume that Theorem 2.114 holds under the condition that $|S| < m$. We must now prove that Theorem 2.114 holds under the condition that $|S| = m$.

We have assumed that Theorem 2.114 holds under the condition that $|S| < m$. In other words, the following claim holds:

> *Claim 1:* Let S be a finite set satisfying $|S| < m$. Let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Then, the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is a 1-element set.

Now, we must now prove that Theorem 2.114 holds under the condition that $|S| = m$. In other words, we must prove the following claim:

> *Claim 2:* Let S be a finite set satisfying $|S| = m$. Let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Then, the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is a 1-element set.

Before we start proving Claim 2, let us prove two auxiliary claims:

> *Claim 3:* Let S be a finite set satisfying $|S| < m$. Let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Let $r \in S$. Let $g \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$ and $c \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{r\}}\right)$. Then, $g = a_r + c$.

[*Proof of Claim 3:* The set $S \setminus \{r\}$ is a subset of the finite set S, and thus itself is finite. Moreover, $r \in S$, so that $|S \setminus \{r\}| = |S| - 1$. Thus, $|S| = |S \setminus \{r\}| + 1$. Hence, the definition of $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ yields

$$\operatorname{Sums}\left((a_s)_{s\in S}\right) = \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)\right\}. \tag{165}$$

But recall that $r \in S$ and $c \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{r\}}\right)$. Hence, the number $a_r + c$ has the form $a_t + b$ for some $t \in S$ and $b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)$ (namely, for $t = r$ and $b = c$). In other words,

$$a_r + c \in \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)\right\}.$$

In view of (165), this rewrites as $a_r + c \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$.

But Claim 1 shows that the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is a 1-element set. Hence, any two elements of $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ are equal. In other words, any $x \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$ and $y \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$ satisfy $x = y$. Applying this to $x = g$ and $y = a_r + c$, we obtain $g = a_r + c$ (since $g \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$ and $a_r + c \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$). This proves Claim 3.]

> *Claim 4:* Let S be a finite set satisfying $|S| = m$. Let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Let $p \in S$ and $q \in S$. Let $f \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{p\}}\right)$ and $g \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{q\}}\right)$. Then, $a_p + f = a_q + g$.

[*Proof of Claim 4:* We have $p \in S$, and thus $|S \setminus \{p\}| = |S| - 1 < |S| = m$. Hence, Claim 1 (applied to $S \setminus \{p\}$ instead of S) yields that the set $\operatorname{Sums}\left((a_s)_{s\in S\setminus\{p\}}\right)$ is a 1-element set. In other words, $\operatorname{Sums}\left((a_s)_{s\in S\setminus\{p\}}\right)$ can be written in the form $\operatorname{Sums}\left((a_s)_{s\in S\setminus\{p\}}\right) = \{h\}$ for some number h. Consider this h.

We are in one of the following two cases:

Case 1: We have $p = q$.

Case 2: We have $p \neq q$.

Let us first consider Case 1. In this case, we have $p = q$. Hence, $q = p$, so that $a_q = a_p$.

We have $f \in \operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p\}}\right) = \{h\}$, so that $f = h$. Also,

$$g \in \operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{q\}}\right) = \operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p\}}\right) \qquad (\text{since } q = p)$$
$$= \{h\},$$

so that $g = h$. Comparing $a_p + \underbrace{f}_{=h} = a_p + h$ with $\underbrace{a_q}_{=a_p} + \underbrace{g}_{=h} = a_p + h$, we obtain $a_p + f = a_q + g$. Hence, Claim 4 is proven in Case 1.

Let us now consider Case 2. In this case, we have $p \neq q$. Thus, $q \neq p$, so that $q \notin \{p\}$.

We have $S \setminus \{p,q\} \subseteq S \setminus \{p\}$ (since $\{p\} \subseteq \{p,q\}$) and thus $|S \setminus \{p,q\}| \leq |S \setminus \{p\}| < m$. Hence, Claim 1 (applied to $S \setminus \{p,q\}$ instead of S) shows that the set $\operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p,q\}}\right)$ is a 1-element set. In other words, $\operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p,q\}}\right)$ has the form

$$\operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p,q\}}\right) = \{c\}$$

for some number c. Consider this c. Hence,

$$c \in \{c\} = \operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p,q\}}\right) = \operatorname{Sums}\left(\left(a_s\right)_{s\in (S\setminus\{p\})\setminus\{q\}}\right)$$

(since $S \setminus \{p,q\} = (S \setminus \{p\}) \setminus \{q\}$). Also, $q \in S \setminus \{p\}$ (since $q \in S$ and $q \notin \{p\}$). Thus, Claim 3 (applied to $S \setminus \{p\}$, q and f instead of S, r and g) yields $f = a_q + c$ (since $|S \setminus \{p\}| < m$ and $f \in \operatorname{Sums}\left(\left(a_s\right)_{s\in S\setminus\{p\}}\right)$ and $c \in \operatorname{Sums}\left(\left(a_s\right)_{s\in (S\setminus\{p\})\setminus\{q\}}\right)$).

The same argument (but with p, q, f and g replaced by q, p, g and f) yields $g = a_p + c$.

Now,

$$a_p + \underbrace{f}_{=a_q+c} = a_p + \left(a_q + c\right) = a_q + \left(a_p + c\right)$$

(by Lemma 2.115, applied to $a = a_p$ and $b = a_q$). Comparing this with

$$a_q + \underbrace{g}_{=a_p+c} = a_q + \left(a_p + c\right),$$

we obtain $a_p + f = a_q + g$. Thus, Claim 4 is proven in Case 2.

We have now proven Claim 4 in both Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Claim 4 always holds.]

We can now prove Claim 2:

[*Proof of Claim 2:* If $|S| = 0$, then Claim 2 holds[79]. Hence, for the rest of this proof of Claim 2, we can WLOG assume that we don't have $|S| = 0$. Assume this.

[79] *Proof.* Assume that $|S| = 0$. Hence, Definition 2.112 yields $\operatorname{Sums}\left(\left(a_s\right)_{s\in S}\right) = \{0\}$. Hence, the set $\operatorname{Sums}\left(\left(a_s\right)_{s\in S}\right)$ is a 1-element set (since the set $\{0\}$ is a 1-element set). In other words, Claim 2 holds. Qed.

We have $|S| \neq 0$ (since we don't have $|S| = 0$). Hence, $|S|$ is a positive integer. Thus, $|S| - 1 \in \mathbb{N}$. Also, the set S is nonempty (since $|S| \neq 0$). Hence, there exists some $p \in S$. Consider this p.

We have $p \in S$ and thus $|S \setminus \{p\}| = |S| - 1 < |S| = m$. Hence, Claim 1 (applied to $S \setminus \{p\}$ instead of S) shows that the set $\operatorname{Sums}\left((a_s)_{s \in S \setminus \{p\}}\right)$ is a 1-element set. In other words, $\operatorname{Sums}\left((a_s)_{s \in S \setminus \{p\}}\right)$ has the form

$$\operatorname{Sums}\left((a_s)_{s \in S \setminus \{p\}}\right) = \{f\}$$

for some number f. Consider this f. Thus,

$$f \in \{f\} = \operatorname{Sums}\left((a_s)_{s \in S \setminus \{p\}}\right). \tag{166}$$

Define $n \in \mathbb{N}$ by $n = |S| - 1$. (This is allowed, since $|S| - 1 \in \mathbb{N}$.) Then, from $n = |S| - 1$, we obtain $|S| = n + 1$. Hence, the definition of $\operatorname{Sums}\left((a_s)_{s \in S}\right)$ yields

$$\operatorname{Sums}\left((a_s)_{s \in S}\right) = \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{t\}}\right)\right\}. \tag{167}$$

But recall that $p \in S$ and $f \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{p\}}\right)$. Hence, the number $a_p + f$ has the form $a_t + b$ for some $t \in S$ and $b \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{t\}}\right)$ (namely, for $t = p$ and $b = f$). In other words,

$$a_p + f \in \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{t\}}\right)\right\}.$$

In view of (167), this rewrites as

$$a_p + f \in \operatorname{Sums}\left((a_s)_{s \in S}\right).$$

Thus,

$$\{a_p + f\} \subseteq \operatorname{Sums}\left((a_s)_{s \in S}\right). \tag{168}$$

Next, we shall show the reverse inclusion (i.e., we shall show that $\operatorname{Sums}\left((a_s)_{s \in S}\right) \subseteq \{a_p + f\}$).

Indeed, let $w \in \operatorname{Sums}\left((a_s)_{s \in S}\right)$. Thus,

$$w \in \operatorname{Sums}\left((a_s)_{s \in S}\right) = \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{t\}}\right)\right\}$$

(by (167)). In other words, w can be written as $w = a_t + b$ for some $t \in S$ and $b \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{t\}}\right)$. Consider these t and b, and denote them by q and g. Thus, $q \in S$ and $g \in \operatorname{Sums}\left((a_s)_{s \in S \setminus \{q\}}\right)$ satisfy $w = a_q + g$.

But Claim 4 yields $a_p + f = a_q + g$. Comparing this with $w = a_q + g$, we obtain $w = a_p + f$. Thus, $w \in \{a_p + f\}$.

Now, forget that we fixed w. We thus have proven that $w \in \{a_p + f\}$ for each $w \in \operatorname{Sums}\left((a_s)_{s\in S}\right)$. In other words, $\operatorname{Sums}\left((a_s)_{s\in S}\right) \subseteq \{a_p + f\}$. Combining this with (168), we conclude that $\operatorname{Sums}\left((a_s)_{s\in S}\right) = \{a_p + f\}$. Hence, the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is a 1-element set. This proves Claim 2.]

Now, we have proven Claim 2. But Claim 2 says precisely that Theorem 2.114 holds under the condition that $|S| = m$. Hence, we have proven that Theorem 2.114 holds under the condition that $|S| = m$. This completes the induction step. Thus, Theorem 2.114 is proven by strong induction. □

2.14.6. Sums of numbers are well-defined

We can now give a new definition of the sum $\sum_{s\in S} a_s$ (for any finite set S and any $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$), which is different from Definition 2.111 and (unlike the latter) is clearly legitimate:

Definition 2.116. Let S be a finite set, and let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Then, the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is a 1-element set (by Theorem 2.114). We define $\sum_{s\in S} a_s$ to be the unique element of this set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$.

However, we have not reached our goal yet: After all, we wanted to prove that Definition 2.111 is legitimate, rather than replace it by a new definition!

Fortunately, we are very close to achieving this goal (after having done all the hard work in the proof of Theorem 2.114 above); we are soon going to show that Definition 2.111 is justified **and** that it is equivalent to Definition 2.116 (that is, both definitions yield the same value of $\sum_{s\in S} a_s$). First, we need a simple lemma, which says that the notation $\sum_{s\in S} a_s$ defined in Definition 2.116 "behaves" like the one defined in Definition 2.111:

Lemma 2.117. In this lemma, we shall use Definition 2.116 (not Definition 2.111).

Let S be a finite set, and let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family.

(a) If $|S| = 0$, then

$$\sum_{s\in S} a_s = 0.$$

(b) For any $t \in S$, we have

$$\sum_{s\in S} a_s = a_t + \sum_{s\in S\setminus\{t\}} a_s.$$

Proof of Lemma 2.117. **(a)** Assume that $|S| = 0$. Thus, $\operatorname{Sums}\left((a_s)_{s\in S}\right) = \{0\}$ (by the definition of $\operatorname{Sums}\left((a_s)_{s\in S}\right)$). Hence, the unique element of the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is 0.

But Definition 2.116 yields that $\sum_{s\in S} a_s$ is the unique element of the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$. Thus, $\sum_{s\in S} a_s$ is 0 (since the unique element of the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ is 0). In other words, $\sum_{s\in S} a_s = 0$. This proves Lemma 2.117 **(a)**.

(b) Let $p \in S$. Thus, $|S \setminus \{p\}| = |S| - 1$, so that $|S| = |S \setminus \{p\}| + 1$. Hence, the definition of $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ yields

$$\operatorname{Sums}\left((a_s)_{s\in S}\right) = \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)\right\}. \tag{169}$$

Definition 2.116 yields that $\sum_{s\in S\setminus\{p\}} a_s$ is the unique element of the set $\operatorname{Sums}\left((a_s)_{s\in S\setminus\{p\}}\right)$. Thus, $\sum_{s\in S\setminus\{p\}} a_s \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{p\}}\right)$. Thus, $a_p + \sum_{s\in S\setminus\{p\}} a_s$ is a number of the form $a_t + b$ for some $t \in S$ and some $b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)$ (namely, for $t = p$ and $b = \sum_{s\in S\setminus\{p\}} a_s$). In other words,

$$a_p + \sum_{s\in S\setminus\{p\}} a_s \in \left\{a_t + b \mid t \in S \text{ and } b \in \operatorname{Sums}\left((a_s)_{s\in S\setminus\{t\}}\right)\right\}.$$

In view of (169), this rewrites as

$$a_p + \sum_{s\in S\setminus\{p\}} a_s \in \operatorname{Sums}\left((a_s)_{s\in S}\right). \tag{170}$$

But Definition 2.116 yields that $\sum_{s\in S} a_s$ is the unique element of the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$. Hence, the set $\operatorname{Sums}\left((a_s)_{s\in S}\right)$ consists only of the element $\sum_{s\in S} a_s$. In other words,

$$\operatorname{Sums}\left((a_s)_{s\in S}\right) = \left\{\sum_{s\in S} a_s\right\}.$$

Thus, (170) rewrites as

$$a_p + \sum_{s\in S\setminus\{p\}} a_s \in \left\{\sum_{s\in S} a_s\right\}.$$

In other words, $a_p + \sum_{s\in S\setminus\{p\}} a_s = \sum_{s\in S} a_s$. Thus, $\sum_{s\in S} a_s = a_p + \sum_{s\in S\setminus\{p\}} a_s$.

Now, forget that we fixed p. We thus have proven that for any $p \in S$, we have $\sum_{s\in S} a_s = a_p + \sum_{s\in S\setminus\{p\}} a_s$. Renaming the variable p as t in this statement, we obtain the following: For any $t \in S$, we have $\sum_{s\in S} a_s = a_t + \sum_{s\in S\setminus\{t\}} a_s$. This proves Lemma 2.117 **(b)**. □

We can now finally state what we wanted to state:

Theorem 2.118. **(a)** Definition 2.111 is legitimate: i.e., the value of $\sum_{s \in S} a_s$ in Definition 2.111 does not depend on the choice of t.

(b) Definition 2.111 is equivalent to Definition 2.116: i.e., both of these definitions yield the same value of $\sum_{s \in S} a_s$.

It makes sense to call Theorem 2.118 **(a)** the *general commutativity theorem*, as it says that a sum of n numbers can be computed in an arbitrary order.

Proof of Theorem 2.118. Let us first use Definition 2.116 (not Definition 2.111). Then, for any finite set S and any $\mathbb{A}$-valued S-family $(a_s)_{s \in S}$, we can compute the number $\sum_{s \in S} a_s$ by the following algorithm (which uses recursion on $|S|$):

- If $|S| = 0$, then $\sum_{s \in S} a_s = 0$. (This follows from Lemma 2.117 **(a)**.)
- Otherwise, we have $|S| = n + 1$ for some $n \in \mathbb{N}$. Consider this n. Thus, $|S| = n + 1 \geq 1 > 0$, so that the set S is nonempty. Fix any $t \in S$. (Such a t exists, since the set S is nonempty.) We have $|S \setminus \{t\}| = |S| - 1 = n$ (since $|S| = n + 1$), so that we can assume (because we are using recursion) that $\sum_{s \in S \setminus \{t\}} a_s$ has already been computed. Then, $\sum_{s \in S} a_s = a_t + \sum_{s \in S \setminus \{t\}} a_s$. (This follows from Lemma 2.117 **(b)**.)

We can restate this algorithm as an alternative definition of $\sum_{s \in S} a_s$; it then takes the following form:

> *Alternative definition of $\sum_{s \in S} a_s$ for any finite set S and any $\mathbb{A}$-valued S-family $(a_s)_{s \in S}$:* If S is a finite set, and if $(a_s)_{s \in S}$ is an $\mathbb{A}$-valued S-family, then we define $\sum_{s \in S} a_s$ by recursion on $|S|$ as follows:
>
> - If $|S| = 0$, then $\sum_{s \in S} a_s$ is defined to be 0.
> - Let $n \in \mathbb{N}$. Assume that we have defined $\sum_{s \in S} a_s$ for every finite set S with $|S| = n$ and any $\mathbb{A}$-valued S-family $(a_s)_{s \in S}$. Now, if S is a finite set with $|S| = n + 1$, and if $(a_s)_{s \in S}$ is any $\mathbb{A}$-valued S-family, then $\sum_{s \in S} a_s$ is defined by picking any $t \in S$ and setting
>
> $$\sum_{s \in S} a_s = a_t + \sum_{s \in S \setminus \{t\}} a_s. \qquad (171)$$

This alternative definition of $\sum_{s \in S} a_s$ merely follows the above algorithm for computing $\sum_{s \in S} a_s$. Thus, it is guaranteed to always yield the same value of $\sum_{s \in S} a_s$ as Definition 2.116, independently of the choice of t. Hence, we obtain the following:

Claim 1: This alternative definition is legitimate (i.e., the value of $\sum_{s\in S} a_s$ in (171) does not depend on the choice of t), and is equivalent to Definition 2.116.

But on the other hand, this alternative definition is precisely Definition 2.111. Hence, Claim 1 rewrites as follows: Definition 2.111 is legitimate (i.e., the value of $\sum_{s\in S} a_s$ in Definition 2.111 does not depend on the choice of t), and is equivalent to Definition 2.116. This proves both parts **(a)** and **(b)** of Theorem 2.118. $\square$

Theorem 2.118 **(a)** shows that Definition 2.111 is legitimate.

Thus, at last, we have vindicated the notation $\sum_{s\in S} a_s$ that was introduced in Section 1.4 (because the definition of this notation we gave in Section 1.4 was precisely Definition 2.111). We can now forget about Definition 2.116, since it has served its purpose (which was to justify Definition 2.111). (Of course, we could also forget about Definition 2.111 instead, and use Definition 2.116 as our definition of $\sum_{s\in S} a_s$ (after all, these two definitions are equivalent, as we now know). Then, we would have to replace every reference to the definition of $\sum_{s\in S} a_s$ by a reference to Lemma 2.117; in particular, we would have to replace every use of (1) by a use of Lemma 2.117 **(b)**. Other than this, everything would work the same way.)

The notation $\sum_{s\in S} a_s$ has several properties, many of which were collected in Section 1.4. We shall prove some of these properties later in this section.

From now on, we shall be using all the conventions and notations regarding sums that we introduced in Section 1.4. In particular, expressions of the form "$\sum_{s\in S} a_s + b$" shall always be interpreted as $\left(\sum_{s\in S} a_s\right) + b$, not as $\sum_{s\in S} (a_s + b)$; but expressions of the form "$\sum_{s\in S} ba_s c$" shall always be understood to mean $\sum_{s\in S} (ba_s c)$.

2.14.7. Triangular numbers revisited

Recall one specific notation we introduced in Section 1.4: If u and v are two integers, and if a_s is a number for each $s \in \{u, u+1, \ldots, v\}$, then $\sum_{s=u}^{v} a_s$ is defined by

$$\sum_{s=u}^{v} a_s = \sum_{s\in\{u,u+1,\ldots,v\}} a_s.$$

This sum $\sum_{s=u}^{v} a_s$ is also denoted by $a_u + a_{u+1} + \cdots + a_v$.

We are now ready to do something that we evaded in Section 2.4: namely, to speak of the sum of the first n positive integers without having to define it recursively. Indeed, we can now interpret this sum as $\sum_{i\in\{1,2,\ldots,n\}} i$, an expression which has a well-defined meaning because we have shown that the notation $\sum_{s\in S} a_s$ is well-defined. We can also rewrite this expression as $\sum_{i=1}^{n} i$ or as $1 + 2 + \cdots + n$.

Thus, the classical fact that the sum of the first n positive integers is $\frac{n(n+1)}{2}$ can now be stated as follows:

Proposition 2.119. We have

$$\sum_{i\in\{1,2,\ldots,n\}} i = \frac{n(n+1)}{2} \qquad \text{for each } n \in \mathbb{N}. \tag{172}$$

Proof of Proposition 2.119. We shall prove (172) by induction on n:

Induction base: We have $\{1, 2, \ldots, 0\} = \varnothing$ and thus $|\{1, 2, \ldots, 0\}| = |\varnothing| = 0$. Hence, the definition of $\sum_{i\in\{1,2,\ldots,0\}} i$ yields

$$\sum_{i\in\{1,2,\ldots,0\}} i = 0. \tag{173}$$

(To be more precise, we have used the first bullet point of Definition 2.111 here, which says that $\sum_{s\in S} a_s = 0$ whenever the set S and the $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$ satisfy $|S| = 0$. If you are using Definition 2.116 instead of Definition 2.111, you should instead be using Lemma 2.117 **(a)** to argue this.)

Comparing (173) with $\frac{0(0+1)}{2} = 0$, we obtain $\sum_{i\in\{1,2,\ldots,0\}} i = \frac{0(0+1)}{2}$. In other words, (172) holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that (172) holds for $n = m$. We must prove that (172) holds for $n = m + 1$.

We have assumed that (172) holds for $n = m$. In other words, we have

$$\sum_{i\in\{1,2,\ldots,m\}} i = \frac{m(m+1)}{2}. \tag{174}$$

Now, $|\{1, 2, \ldots, m+1\}| = m + 1$ and $m + 1 \in \{1, 2, \ldots, m+1\}$ (since $m + 1$ is a positive integer (since $m \in \mathbb{N}$)). Hence, (1) (applied to $n = m$, $S = \{1, 2, \ldots, m+1\}$, $t = m + 1$ and $(a_s)_{s\in S} = (i)_{i\in\{1,2,\ldots,m+1\}}$) yields

$$\sum_{i\in\{1,2,\ldots,m+1\}} i = (m+1) + \sum_{i\in\{1,2,\ldots,m+1\}\setminus\{m+1\}} i. \tag{175}$$

(Here, we have relied on the equality (1), which appears verbatim in Definition 2.111. If you are using Definition 2.116 instead of Definition 2.111, you should instead be using Lemma 2.117 **(b)** to argue this.)

Now, (175) becomes

$$\sum_{i\in\{1,2,\ldots,m+1\}} i = (m+1) + \sum_{i\in\{1,2,\ldots,m+1\}\setminus\{m+1\}} i = (m+1) + \underbrace{\sum_{i\in\{1,2,\ldots,m\}} i}_{\substack{=\dfrac{m(m+1)}{2} \\ \text{(by (174))}}}$$

$$\left(\text{since } \{1,2,\ldots,m+1\}\setminus\{m+1\} = \{1,2,\ldots,m\}\right)$$

$$= (m+1) + \frac{m(m+1)}{2} = \frac{2(m+1)+m(m+1)}{2}$$

$$= \frac{(m+1)((m+1)+1)}{2}$$

(since $2(m+1)+m(m+1) = (m+1)((m+1)+1)$). In other words, (172) holds for $n = m+1$. This completes the induction step. Thus, the induction proof of (172) is finished. Hence, Proposition 2.119 holds. □

2.14.8. Sums of a few numbers

Merely for the sake of future convenience, let us restate (1) in a slightly more direct way (without mentioning $|S|$):

Proposition 2.120. Let S be a finite set, and let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Let $t \in S$. Then,

$$\sum_{s\in S} a_s = a_t + \sum_{s\in S\setminus\{t\}} a_s.$$

Proof of Proposition 2.120. Let $n = |S \setminus \{t\}|$; thus, $n \in \mathbb{N}$ (since $S\setminus\{t\}$ is a finite set). Also, $n = |S\setminus\{t\}| = |S| - 1$ (since $t \in S$), and thus $|S| = n+1$. Hence, (1) yields $\sum_{s\in S} a_s = a_t + \sum_{s\in S\setminus\{t\}} a_s$. This proves Proposition 2.120. □

(Alternatively, we can argue that Proposition 2.120 is the same as Lemma 2.117 **(b)**, except that we are now using Definition 2.111 instead of Definition 2.116 to define the sums involved – but this difference is insubstantial, since we have shown that these two definitions are equivalent.)

In Section 1.4, we have introduced $a_u + a_{u+1} + \cdots + a_v$ as an abbreviation for the sum $\sum_{s=u}^{v} a_s = \sum_{s\in\{u,u+1,\ldots,v\}} a_s$ (whenever u and v are two integers, and a_s is a number for each $s \in \{u, u+1, \ldots, v\}$). In order to ensure that this abbreviation does not create any nasty surprises, we need to check that it behaves as we would expect – i.e., that it satisfies the following four properties:

- If the sum $a_u + a_{u+1} + \cdots + a_v$ has no addends (i.e., if $u > v$), then it equals 0.

- If the sum $a_u + a_{u+1} + \cdots + a_v$ has exactly one addend (i.e., if $u = v$), then it equals a_u.
- If the sum $a_u + a_{u+1} + \cdots + a_v$ has exactly two addends (i.e., if $u = v - 1$), then it equals $a_u + a_v$.
- If $v \geq u$, then

$$a_u + a_{u+1} + \cdots + a_v = \left(a_u + a_{u+1} + \cdots + a_{v-1}\right) + a_v$$
$$= a_u + \left(a_{u+1} + a_{u+2} + \cdots + a_v\right).$$

The first of these four properties follows from the definition (indeed, if $u > v$, then the set $\{u, u+1, \ldots, v\}$ is empty and thus satisfies $|\{u, u+1, \ldots, v\}| = 0$; but this yields $\sum_{s \in \{u,u+1,\ldots,v\}} a_s = 0$). The fourth of these four properties can easily be obtained from Proposition 2.120[80]. The second and third properties follow from the following fact:

[80] In more detail: Assume that $v \geq u$. Thus, both u and v belong to the set $\{u, u+1, \ldots, v\}$. Hence, Proposition 2.120 (applied to $S = \{u, u+1, \ldots, v\}$ and $t = v$) yields $\sum_{s \in \{u,u+1,\ldots,v\}} a_s = a_v + \sum_{s \in \{u,u+1,\ldots,v\} \setminus \{v\}} a_s$. Thus,

$$\begin{aligned} & a_u + a_{u+1} + \cdots + a_v \\ &= \sum_{s \in \{u,u+1,\ldots,v\}} a_s = a_v + \sum_{s \in \{u,u+1,\ldots,v\} \setminus \{v\}} a_s \\ &= a_v + \sum_{s \in \{u,u+1,\ldots,v-1\}} a_s \qquad \left(\text{since } \{u, u+1, \ldots, v\} \setminus \{v\} = \{u, u+1, \ldots, v-1\}\right) \\ &= a_v + \left(a_u + a_{u+1} + \cdots + a_{v-1}\right) = \left(a_u + a_{u+1} + \cdots + a_{v-1}\right) + a_v. \end{aligned}$$

Also, Proposition 2.120 (applied to $S = \{u, u+1, \ldots, v\}$ and $t = u$) yields $\sum_{s \in \{u,u+1,\ldots,v\}} a_s = a_u + \sum_{s \in \{u,u+1,\ldots,v\} \setminus \{u\}} a_s$. Thus,

$$\begin{aligned} & a_u + a_{u+1} + \cdots + a_v \\ &= \sum_{s \in \{u,u+1,\ldots,v\}} a_s = a_u + \sum_{s \in \{u,u+1,\ldots,v\} \setminus \{u\}} a_s \\ &= a_u + \sum_{s \in \{u+1,u+2,\ldots,v\}} a_s \qquad \left(\text{since } \{u, u+1, \ldots, v\} \setminus \{u\} = \{u+1, u+2, \ldots, v\}\right) \\ &= a_u + \left(a_{u+1} + a_{u+2} + \cdots + a_v\right). \end{aligned}$$

Hence,

$$a_u + a_{u+1} + \cdots + a_v = \left(a_u + a_{u+1} + \cdots + a_{v-1}\right) + a_v = a_u + \left(a_{u+1} + a_{u+2} + \cdots + a_v\right).$$

Proposition 2.121. Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{A}$.

(a) If $S = \{p\}$ for some element p, then this p satisfies

$$\sum_{s \in S} a_s = a_p.$$

(b) If $S = \{p, q\}$ for two distinct elements p and q, then these p and q satisfy

$$\sum_{s \in S} a_s = a_p + a_q.$$

Proof of Proposition 2.121. **(a)** Assume that $S = \{p\}$ for some element p. Consider this p.

The first bullet point of Definition 2.111 shows that $\sum_{s \in \varnothing} a_s = 0$ (since $|\varnothing| = 0$). But $p \in \{p\} = S$. Hence, Proposition 2.120 (applied to $t = p$) yields

$$\sum_{s \in S} a_s = a_p + \sum_{s \in S \setminus \{p\}} a_s = a_p + \underbrace{\sum_{s \in \varnothing} a_s}_{=0} \qquad \left(\text{since } \underbrace{S}_{=\{p\}} \setminus \{p\} = \{p\} \setminus \{p\} = \varnothing \right)$$
$$= a_p + 0 = a_p.$$

This proves Proposition 2.121 **(a)**.

(b) Assume that $S = \{p, q\}$ for two distinct elements p and q. Consider these p and q. Thus, $q \neq p$ (since p and q are distinct), so that $q \notin \{p\}$.

Proposition 2.121 **(a)** (applied to $\{p\}$ instead of S) yields $\sum_{s \in \{p\}} a_s = a_p$ (since $\{p\} = \{p\}$).

We have $\underbrace{S}_{=\{p,q\}=\{p\} \cup \{q\}} \setminus \{q\} = (\{p\} \cup \{q\}) \setminus \{q\} = \{p\} \setminus \{q\} = \{p\}$ (since $q \notin \{p\}$). Also, $q \in \{p, q\} = S$. Hence, Proposition 2.120 (applied to $t = q$) yields

$$\sum_{s \in S} a_s = a_q + \sum_{s \in S \setminus \{q\}} a_s = a_q + \underbrace{\sum_{s \in \{p\}} a_s}_{=a_p} \qquad (\text{since } S \setminus \{q\} = \{p\})$$
$$= a_q + a_p = a_p + a_q.$$

This proves Proposition 2.121 **(b)**. □

2.14.9. Linearity of sums

We shall now prove some general properties of finite sums. We begin with the equality (7) from Section 1.4:

Theorem 2.122. Let S be a finite set. For every $s \in S$, let a_s and b_s be elements of $\mathbb{A}$. Then,

$$\sum_{s\in S}(a_s+b_s)=\sum_{s\in S}a_s+\sum_{s\in S}b_s.$$

Before we prove this theorem, let us show a simple lemma:

Lemma 2.123. Let x, y, u and v be four numbers (i.e., elements of $\mathbb{A}$). Then,

$$(x+y)+(u+v)=(x+u)+(y+v).$$

Proof of Lemma 2.123. Proposition 2.103 (applied to $a=y$, $b=u$ and $c=v$) yields

$$(y+u)+v=y+(u+v). \tag{176}$$

Also, Proposition 2.103 (applied to $a=x$, $b=y$ and $c=u+v$) yields

$$\begin{aligned}(x+y)+(u+v)&=x+\underbrace{(y+(u+v))}_{\substack{=(y+u)+v\\ \text{(by (176))}}}\\&=x+((y+u)+v).\end{aligned} \tag{177}$$

The same argument (with y and u replaced by u and y) yields

$$(x+u)+(y+v)=x+((u+y)+v). \tag{178}$$

But Proposition 2.104 (applied to $a=y$ and $b=u$) yields $y+u=u+y$. Thus, (177) becomes

$$(x+y)+(u+v)=x+\left(\underbrace{(y+u)}_{=u+y}+v\right)=x+((u+y)+v)=(x+u)+(y+v)$$

(by (178)). This proves Lemma 2.123. □

Proof of Theorem 2.122. Forget that we fixed S, a_s and b_s. We shall prove Theorem 2.122 by induction on $|S|$:

Induction base: Theorem 2.122 holds under the condition that $|S|=0$ [81]. This completes the induction base.

[81] *Proof.* Let S, a_s and b_s be as in Theorem 2.122. Assume that $|S|=0$. Thus, the first bullet point of Definition 2.111 yields $\sum_{s\in S}a_s=0$ and $\sum_{s\in S}b_s=0$ and $\sum_{s\in S}(a_s+b_s)=0$. Hence,

$$\sum_{s\in S}(a_s+b_s)=0=\underbrace{0}_{=\sum_{s\in S}a_s}+\underbrace{0}_{=\sum_{s\in S}b_s}=\sum_{s\in S}a_s+\sum_{s\in S}b_s.$$

Now, forget that we fixed S, a_s and b_s. We thus have proved that if S, a_s and b_s are as in Theorem 2.122, and if $|S|=0$, then $\sum_{s\in S}(a_s+b_s)=\sum_{s\in S}a_s+\sum_{s\in S}b_s$. In other words, Theorem 2.122 holds under the condition that $|S|=0$. Qed.

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.122 holds under the condition that $|S| = m$. We must now prove that Theorem 2.122 holds under the condition that $|S| = m + 1$.

We have assumed that Theorem 2.122 holds under the condition that $|S| = m$. In other words, the following claim holds:

> *Claim 1:* Let S be a finite set such that $|S| = m$. For every $s \in S$, let a_s and b_s be elements of $\mathbb{A}$. Then,
>
> $$\sum_{s \in S} (a_s + b_s) = \sum_{s \in S} a_s + \sum_{s \in S} b_s.$$

Next, we shall show the following claim:

> *Claim 2:* Let S be a finite set such that $|S| = m + 1$. For every $s \in S$, let a_s and b_s be elements of $\mathbb{A}$. Then,
>
> $$\sum_{s \in S} (a_s + b_s) = \sum_{s \in S} a_s + \sum_{s \in S} b_s.$$

[*Proof of Claim 2:* We have $|S| = m + 1 > m \geq 0$. Hence, the set S is nonempty. Thus, there exists some $t \in S$. Consider this t.

From $t \in S$, we obtain $|S \setminus \{t\}| = |S| - 1 = m$ (since $|S| = m + 1$). Hence, Claim 1 (applied to $S \setminus \{t\}$ instead of S) yields

$$\sum_{s \in S \setminus \{t\}} (a_s + b_s) = \sum_{s \in S \setminus \{t\}} a_s + \sum_{s \in S \setminus \{t\}} b_s. \tag{179}$$

Now, Proposition 2.120 yields

$$\sum_{s \in S} a_s = a_t + \sum_{s \in S \setminus \{t\}} a_s. \tag{180}$$

Also, Proposition 2.120 (applied to b_s instead of a_s) yields

$$\sum_{s \in S} b_s = b_t + \sum_{s \in S \setminus \{t\}} b_s. \tag{181}$$

Finally, Proposition 2.120 (applied to $a_s + b_s$ instead of a_s) yields

$$\sum_{s\in S}(a_s+b_s) = (a_t+b_t) + \underbrace{\sum_{s\in S\setminus\{t\}}(a_s+b_s)}_{\substack{=\sum_{s\in S\setminus\{t\}} a_s + \sum_{s\in S\setminus\{t\}} b_s \\ \text{(by (179))}}}$$
$$= (a_t+b_t) + \left(\sum_{s\in S\setminus\{t\}} a_s + \sum_{s\in S\setminus\{t\}} b_s\right)$$
$$= \underbrace{\left(a_t + \sum_{s\in S\setminus\{t\}} a_s\right)}_{\substack{=\sum_{s\in S} a_s \\ \text{(by (180))}}} + \underbrace{\left(b_t + \sum_{s\in S\setminus\{t\}} b_s\right)}_{\substack{=\sum_{s\in S} b_s \\ \text{(by (181))}}}$$
$$\left(\begin{array}{c}\text{by Lemma 2.123 (applied} \\ \text{to } x = a_t,\ y = b_t,\ u = \sum_{s\in S\setminus\{t\}} a_s \text{ and } v = \sum_{s\in S\setminus\{t\}} b_s)\end{array}\right)$$
$$= \sum_{s\in S} a_s + \sum_{s\in S} b_s.$$

This proves Claim 2.]

But Claim 2 says precisely that Theorem 2.122 holds under the condition that $|S| = m+1$. Hence, we conclude that Theorem 2.122 holds under the condition that $|S| = m+1$ (since Claim 2 is proven). This completes the induction step. Thus, Theorem 2.122 is proven by induction. □

We shall next prove (9):

Theorem 2.124. Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{A}$. Also, let λ be an element of $\mathbb{A}$. Then,

$$\sum_{s\in S} \lambda a_s = \lambda \sum_{s\in S} a_s.$$

To prove this theorem, we need the following fundamental fact of arithmetic:

Proposition 2.125. Let x, y and z be three numbers (i.e., elements of $\mathbb{A}$). Then, $x(y+z) = xy + xz$.

Proposition 2.125 is known as the *distributivity* (or *left distributivity*) in $\mathbb{A}$. It is a fundamental result, and its proof can be found in standard textbooks[82].

[82] For example, Proposition 2.125 is proven in [Swanso18, Theorem 3.2.3 (6)] for the case when $\mathbb{A} = \mathbb{N}$; in [Swanso18, Theorem 3.5.4 (6)] for the case when $\mathbb{A} = \mathbb{Z}$; in [Swanso18, Theorem 3.6.4 (6)] for the case when $\mathbb{A} = \mathbb{Q}$; in [Swanso18, Theorem 3.7.13] for the case when $\mathbb{A} = \mathbb{R}$; in [Swanso18, Theorem 3.9.3] for the case when $\mathbb{A} = \mathbb{C}$.

Proof of Theorem 2.124. Forget that we fixed S, a_s and λ. We shall prove Theorem 2.124 by induction on $|S|$:

Induction base: The induction base (i.e., proving that Theorem 2.124 holds under the condition that $|S| = 0$) is similar to the induction base in the proof of Theorem 2.122 above; we thus leave it to the reader.

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.124 holds under the condition that $|S| = m$. We must now prove that Theorem 2.124 holds under the condition that $|S| = m + 1$.

We have assumed that Theorem 2.124 holds under the condition that $|S| = m$. In other words, the following claim holds:

> *Claim 1:* Let S be a finite set such that $|S| = m$. For every $s \in S$, let a_s be an element of $\mathbb{A}$. Also, let λ be an element of $\mathbb{A}$. Then,
> $$\sum_{s \in S} \lambda a_s = \lambda \sum_{s \in S} a_s.$$

Next, we shall show the following claim:

> *Claim 2:* Let S be a finite set such that $|S| = m + 1$. For every $s \in S$, let a_s be an element of $\mathbb{A}$. Also, let λ be an element of $\mathbb{A}$. Then,
> $$\sum_{s \in S} \lambda a_s = \lambda \sum_{s \in S} a_s.$$

[*Proof of Claim 2:* We have $|S| = m + 1 > m \geq 0$. Hence, the set S is nonempty. Thus, there exists some $t \in S$. Consider this t.

From $t \in S$, we obtain $|S \setminus \{t\}| = |S| - 1 = m$ (since $|S| = m + 1$). Hence, Claim 1 (applied to $S \setminus \{t\}$ instead of S) yields

$$\sum_{s \in S \setminus \{t\}} \lambda a_s = \lambda \sum_{s \in S \setminus \{t\}} a_s. \tag{182}$$

Now, Proposition 2.120 yields

$$\sum_{s \in S} a_s = a_t + \sum_{s \in S \setminus \{t\}} a_s.$$

Multiplying both sides of this equality by λ, we obtain

$$\lambda \sum_{s \in S} a_s = \lambda \left(a_t + \sum_{s \in S \setminus \{t\}} a_s \right) = \lambda a_t + \lambda \sum_{s \in S \setminus \{t\}} a_s$$

(by Proposition 2.125 (applied to $x = \lambda$, $y = a_t$ and $z = \sum\limits_{s \in S \setminus \{t\}} a_s$)). Also, Proposition 2.120 (applied to λa_s instead of a_s) yields

$$\sum_{s \in S} \lambda a_s = \lambda a_t + \underbrace{\sum_{s \in S \setminus \{t\}} \lambda a_s}_{\substack{= \lambda \sum\limits_{s \in S \setminus \{t\}} a_s \\ \text{(by (182))}}} = \lambda a_t + \lambda \sum_{s \in S \setminus \{t\}} a_s.$$

Comparing the preceding two equalities, we find

$$\sum_{s\in S} \lambda a_s = \lambda \sum_{s\in S} a_s.$$

This proves Claim 2.]

But Claim 2 says precisely that Theorem 2.124 holds under the condition that $|S| = m+1$. Hence, we conclude that Theorem 2.124 holds under the condition that $|S| = m+1$ (since Claim 2 is proven). This completes the induction step. Thus, Theorem 2.124 is proven by induction. □

Finally, let us prove (10):

Theorem 2.126. Let S be a finite set. Then,

$$\sum_{s\in S} 0 = 0.$$

Proof of Theorem 2.126. It is completely straightforward to prove Theorem 2.126 by induction on $|S|$ (as we proved Theorem 2.124, for example). But let us give an even shorter argument: Theorem 2.124 (applied to $a_s = 0$ and $\lambda = 0$) yields

$$\sum_{s\in S} 0 \cdot 0 = 0 \sum_{s\in S} 0 = 0.$$

In view of $0 \cdot 0 = 0$, this rewrites as $\sum_{s\in S} 0 = 0$. This proves Theorem 2.126. □

2.14.10. Splitting a sum by a value of a function

We shall now prove a more complicated (but crucial) property of finite sums – namely, the equality (22) in the case when W is finite[83]:

Theorem 2.127. Let S be a finite set. Let W be a finite set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s\in S} a_s = \sum_{w\in W} \sum_{\substack{s\in S;\\ f(s)=w}} a_s.$$

Here, we are using the following convention (made in Section 1.4):

[83] We prefer to only treat the case when W is finite for now. The case when W is infinite would require us to properly introduce the notion of an infinite sum with only finitely many nonzero terms. While this is not hard to do, we aren't quite ready for it yet (see Theorem 2.147 further below for this).

Convention 2.128. Let S be a finite set. Let $\mathcal{A}(s)$ be a logical statement defined for every $s \in S$. For each $s \in S$ satisfying $\mathcal{A}(s)$, let a_s be a number (i.e., an element of $\mathbb{A}$). Then, we set

$$\sum_{\substack{s \in S; \\ \mathcal{A}(s)}} a_s = \sum_{s \in \{t \in S \mid \mathcal{A}(t)\}} a_s .$$

Thus, the sum $\sum_{\substack{s \in S; \\ f(s)=w}} a_s$ in Theorem 2.127 can be rewritten as $\sum_{s \in \{t \in S \mid f(t)=w\}} a_s$.

Our proof of Theorem 2.127 relies on the following simple set-theoretic fact:

Lemma 2.129. Let S and W be two sets. Let $f : S \to W$ be a map. Let $q \in S$. Let g be the restriction $f\mid_{S \setminus \{q\}}$ of the map f to $S \setminus \{q\}$. Let $w \in W$. Then,

$$\{t \in S \setminus \{q\} \mid g(t) = w\} = \{t \in S \mid f(t) = w\} \setminus \{q\} .$$

Proof of Lemma 2.129. We know that g is the restriction $f\mid_{S \setminus \{q\}}$ of the map f to $S \setminus \{q\}$. Thus, g is a map from $S \setminus \{q\}$ to W and satisfies

$$g(t) = f(t) \qquad \text{for each } t \in S \setminus \{q\} . \tag{183}$$

Now,

$$\begin{aligned} &\left\{ t \in S \setminus \{q\} \mid \underbrace{g(t)}_{\substack{=f(t) \\ \text{(by (183))}}} = w \right\} \\ &= \{t \in S \setminus \{q\} \mid f(t) = w\} = \{t \in S \mid f(t) = w \text{ and } t \in S \setminus \{q\}\} \\ &= \{t \in S \mid f(t) = w\} \cap \underbrace{\{t \in S \mid t \in S \setminus \{q\}\}}_{=S \setminus \{q\}} \\ &= \{t \in S \mid f(t) = w\} \cap (S \setminus \{q\}) = \{t \in S \mid f(t) = w\} \setminus \{q\} . \end{aligned}$$

This proves Lemma 2.129. ☐

Proof of Theorem 2.127. We shall prove Theorem 2.127 by induction on $|S|$:

Induction base: Theorem 2.127 holds under the condition that $|S| = 0$ [84]. This completes the induction base.

[84] *Proof.* Let S, W, f and a_s be as in Theorem 2.127. Assume that $|S| = 0$. Thus, the first bullet point of Definition 2.111 yields $\sum_{s \in S} a_s = 0$. Moreover, $S = \varnothing$ (since $|S| = 0$). Hence, each $w \in W$

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.127 holds under the condition that $|S| = m$. We must now prove that Theorem 2.127 holds under the condition that $|S| = m + 1$.

We have assumed that Theorem 2.127 holds under the condition that $|S| = m$. In other words, the following claim holds:

Claim 1: Let S be a finite set such that $|S| = m$. Let W be a finite set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s \in S} a_s = \sum_{w \in W} \sum_{\substack{s \in S; \\ f(s) = w}} a_s.$$

Next, we shall show the following claim:

Claim 2: Let S be a finite set such that $|S| = m + 1$. Let W be a finite set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\sum_{s \in S} a_s = \sum_{w \in W} \sum_{\substack{s \in S; \\ f(s) = w}} a_s.$$

[*Proof of Claim 2:* We have $|S| = m + 1 > m \geq 0$. Hence, the set S is nonempty. Thus, there exists some $q \in S$. Consider this q.

From $q \in S$, we obtain $|S \setminus \{q\}| = |S| - 1 = m$ (since $|S| = m + 1$).

Let g be the restriction $f\mid_{S \setminus \{q\}}$ of the map f to $S \setminus \{q\}$. Thus, g is a map from $S \setminus \{q\}$ to W.

---

satisfies

$$\sum_{\substack{s \in S; \\ f(s) = w}} a_s = \sum_{s \in \{t \in S \mid f(t) = w\}} a_s = \sum_{s \in \varnothing} a_s \qquad \left(\begin{array}{c} \text{since } \{t \in S \mid f(s) = w\} = \varnothing \\ \text{(because } \{t \in S \mid f(s) = w\} \subseteq S = \varnothing) \end{array} \right)$$
$$= (\text{empty sum}) = 0.$$

Summing these equalities over all $w \in W$, we obtain

$$\sum_{w \in W} \sum_{\substack{s \in S; \\ f(s) = w}} a_s = \sum_{w \in W} 0 = 0$$

(by an application of Theorem 2.126). Comparing this with $\sum_{s \in S} a_s = 0$, we obtain $\sum_{s \in S} a_s = \sum_{w \in W} \sum_{\substack{s \in S; \\ f(s) = w}} a_s$.

Now, forget that we fixed S, W, f and a_s. We thus have proved that if S, W, f and a_s are as in Theorem 2.127, and if $|S| = 0$, then $\sum_{s \in S} a_s = \sum_{w \in W} \sum_{\substack{s \in S; \\ f(s) = w}} a_s$. In other words, Theorem 2.127 holds under the condition that $|S| = 0$. Qed.

For each $w \in W$, we define a number b_w by

$$b_w = \sum_{\substack{s \in S;\\ f(s)=w}} a_s. \tag{184}$$

Furthermore, for each $w \in W$, we define a number c_w by

$$c_w = \sum_{\substack{s \in S \setminus \{q\};\\ g(s)=w}} a_s. \tag{185}$$

Recall that $|S \setminus \{q\}| = m$. Hence, Claim 1 (applied to $S \setminus \{q\}$ and g instead of S and f) yields

$$\sum_{s \in S \setminus \{q\}} a_s = \sum_{w \in W} \underbrace{\sum_{\substack{s \in S \setminus \{q\};\\ g(s)=w}} a_s}_{\substack{=c_w \\ \text{(by (185))}}} = \sum_{w \in W} c_w. \tag{186}$$

Every $w \in W \setminus \{f(q)\}$ satisfies

$$b_w = c_w. \tag{187}$$

[*Proof of (187):* Let $w \in W \setminus \{f(q)\}$. Thus, $w \in W$ and $w \notin \{f(q)\}$.

If we had $q \in \{t \in S \mid f(t) = w\}$, then we would have $f(q) = w$, which would lead to $w = f(q) \in \{f(q)\}$; but this would contradict $w \notin \{f(q)\}$. Hence, we cannot have $q \in \{t \in S \mid f(t) = w\}$. Hence, we have $q \notin \{t \in S \mid f(t) = w\}$.

But $w \in W$; thus, Lemma 2.129 yields

$$\begin{aligned} \{t \in S \setminus \{q\} \mid g(t) = w\} &= \{t \in S \mid f(t) = w\} \setminus \{q\} \\ &= \{t \in S \mid f(t) = w\} \end{aligned} \tag{188}$$

(since $q \notin \{t \in S \mid f(t) = w\}$).

On the other hand, the definition of b_w yields

$$b_w = \sum_{\substack{s \in S;\\ f(s)=w}} a_s = \sum_{s \in \{t \in S \mid f(t)=w\}} a_s \tag{189}$$

(by the definition of the "$\sum_{\substack{s \in S;\\ f(s)=w}}$" symbol). Also, the definition of c_w yields

$$\begin{aligned} c_w &= \sum_{\substack{s \in S \setminus \{q\};\\ g(s)=w}} a_s = \sum_{s \in \{t \in S \setminus \{q\} \mid g(t)=w\}} a_s = \sum_{s \in \{t \in S \mid f(t)=w\}} a_s \\ &\qquad \left(\text{since } \{t \in S \setminus \{q\} \mid g(t) = w\} = \{t \in S \mid f(t) = w\} \text{ (by (188))} \right) \\ &= b_w \qquad \text{(by (189))}. \end{aligned}$$

Thus, $b_w = c_w$. This proves (187).]

Also,

$$b_{f(q)} = a_q + c_{f(q)}. \tag{190}$$

[*Proof of (190):* Define a subset U of S by

$$U = \{t \in S \mid f(t) = f(q)\}. \tag{191}$$

We can apply Lemma 2.129 to $w = f(q)$. We thus obtain

$$\begin{aligned}\{t \in S \setminus \{q\} \mid g(t) = f(q)\} &= \underbrace{\{t \in S \mid f(t) = f(q)\}}_{\substack{=U \\ \text{(by (191))}}} \setminus \{q\} \\ &= U \setminus \{q\}. \end{aligned} \tag{192}$$

We know that q is a $t \in S$ satisfying $f(t) = f(q)$ (since $q \in S$ and $f(q) = f(q)$). In other words, $q \in \{t \in S \mid f(t) = f(q)\}$. In other words, $q \in U$ (since $U = \{t \in S \mid f(t) = f(q)\}$). Thus, Proposition 2.120 (applied to U and q instead of S and t) yields

$$\sum_{s \in U} a_s = a_q + \sum_{s \in U \setminus \{q\}} a_s. \tag{193}$$

But (192) shows that $U \setminus \{q\} = \{t \in S \setminus \{q\} \mid g(t) = f(q)\}$. Thus,

$$\sum_{s \in U \setminus \{q\}} a_s = \sum_{s \in \{t \in S \setminus \{q\} \mid g(t) = f(q)\}} a_s = c_{f(q)} \tag{194}$$

(since the definition of $c_{f(q)}$ yields $c_{f(q)} = \sum_{\substack{s \in S \setminus \{q\}; \\ g(s) = f(q)}} a_s = \sum_{s \in \{t \in S \setminus \{q\} \mid g(t) = f(q)\}} a_s$).

On the other hand, the definition of $b_{f(q)}$ yields

$$\begin{aligned} b_{f(q)} &= \sum_{\substack{s \in S; \\ f(s) = f(q)}} a_s = \sum_{s \in \{t \in S \mid f(t) = f(q)\}} a_s \\ &= \sum_{s \in U} a_s \qquad \left(\text{since } \{t \in S \mid f(t) = f(q)\} = U\right) \\ &= a_q + \underbrace{\sum_{s \in U \setminus \{q\}} a_s}_{\substack{= c_{f(q)} \\ \text{(by (194))}}} \qquad \text{(by (193))} \\ &= a_q + c_{f(q)}. \end{aligned}$$

This proves (190).]

Now, recall that $q \in S$. Hence, Proposition 2.120 (applied to $t = q$) yields

$$\sum_{s \in S} a_s = a_q + \sum_{s \in S \setminus \{q\}} a_s. \tag{195}$$

Also, $f(q) \in W$. Hence, Proposition 2.120 (applied to W, $(c_w)_{w\in W}$ and $f(q)$ instead of S, $(a_s)_{s\in S}$ and t) yields

$$\sum_{w\in W} c_w = c_{f(q)} + \sum_{w\in W\setminus\{f(q)\}} c_w.$$

Hence, (186) becomes

$$\sum_{s\in S\setminus\{q\}} a_s = \sum_{w\in W} c_w = c_{f(q)} + \sum_{w\in W\setminus\{f(q)\}} c_w. \tag{196}$$

Also, Proposition 2.120 (applied to W, $(b_w)_{w\in W}$ and $f(q)$ instead of S, $(a_s)_{s\in S}$ and t) yields

$$\begin{aligned}\sum_{w\in W} b_w &= \underbrace{b_{f(q)}}_{\substack{=a_q+c_{f(q)} \\ \text{(by (190))}}} + \sum_{w\in W\setminus\{f(q)\}} \underbrace{b_w}_{\substack{=c_w \\ \text{(by (187))}}} \\ &= \left(a_q + c_{f(q)}\right) + \sum_{w\in W\setminus\{f(q)\}} c_w = a_q + \left(c_{f(q)} + \sum_{w\in W\setminus\{f(q)\}} c_w\right)\end{aligned}$$

(by Proposition 2.103, applied to a_q, $c_{f(q)}$ and $\sum\limits_{w\in W\setminus\{f(q)\}} c_w$ instead of a, b and c). Thus,

$$\sum_{w\in W} b_w = a_q + \underbrace{\left(c_{f(q)} + \sum_{w\in W\setminus\{f(q)\}} c_w\right)}_{\substack{=\sum\limits_{s\in S\setminus\{q\}} a_s \\ \text{(by (196))}}} = a_q + \sum_{s\in S\setminus\{q\}} a_s = \sum_{s\in S} a_s$$

(by (195)). Hence,

$$\sum_{s\in S} a_s = \sum_{w\in W} \underbrace{b_w}_{\substack{=\sum\limits_{\substack{s\in S;\\ f(s)=w}} a_s \\ \text{(by (184))}}} = \sum_{w\in W} \sum_{\substack{s\in S;\\ f(s)=w}} a_s.$$

This proves Claim 2.]

But Claim 2 says precisely that Theorem 2.127 holds under the condition that $|S| = m+1$. Hence, we conclude that Theorem 2.127 holds under the condition that $|S| = m+1$ (since Claim 2 is proven). This completes the induction step. Thus, Theorem 2.127 is proven by induction. □

2.14.11. Splitting a sum into two

Next, we shall prove the equality (3):

Theorem 2.130. Let S be a finite set. Let X and Y be two subsets of S such that $X \cap Y = \varnothing$ and $X \cup Y = S$. (Equivalently, X and Y are two subsets of S such that each element of S lies in **exactly** one of X and Y.) Let a_s be a number (i.e., an element of $\mathbb{A}$) for each $s \in S$. Then,

$$\sum_{s\in S} a_s = \sum_{s\in X} a_s + \sum_{s\in Y} a_s.$$

Proof of Theorem 2.130. From the assumptions $X \cap Y = \varnothing$ and $X \cup Y = S$, we can easily obtain $S \setminus X = Y$.

We define a map $f : S \to \{0,1\}$ by setting

$$\left(f(s) = \begin{cases} 0, & \text{if } s \in X; \\ 1, & \text{if } s \notin X \end{cases} \qquad \text{for every } s \in S \right).$$

For each $w \in \{0,1\}$, we define a number b_w by

$$b_w = \sum_{\substack{s\in S;\\ f(s)=w}} a_s. \tag{197}$$

Proposition 2.121 **(b)** (applied to $\{0,1\}$, 0, 1 and $(b_w)_{w\in\{0,1\}}$ instead of S, p, q and $(a_s)_{s\in S}$) yields $\sum_{w\in\{0,1\}} b_w = b_0 + b_1$.

Now, Theorem 2.127 (applied to $W = \{0,1\}$) yields

$$\sum_{s\in S} a_s = \sum_{w\in\{0,1\}} \underbrace{\sum_{\substack{s\in S;\\ f(s)=w}} a_s}_{\substack{=b_w\\ \text{(by (197))}}} = \sum_{w\in\{0,1\}} b_w = b_0 + b_1. \tag{198}$$

On the other hand,

$$b_0 = \sum_{s\in X} a_s. \tag{199}$$

[*Proof of (199):* The definition of the map f shows that an element $t \in S$ satisfies $f(t) = 0$ **if and only if** it belongs to X. Hence, the set of all elements $t \in S$ that satisfy $f(t) = 0$ is precisely X. In other words,

$$\{t \in S \mid f(t) = 0\} = X.$$

But the definition of b_0 yields

$$b_0 = \sum_{\substack{s\in S;\\ f(s)=0}} a_s = \sum_{s\in\{t\in S \mid f(t)=0\}} a_s = \sum_{s\in X} a_s$$

(since $\{t \in S \mid f(t) = 0\} = X$). This proves (199).]

Furthermore,

$$b_1 = \sum_{s \in Y} a_s. \qquad (200)$$

[*Proof of (200):* The definition of the map f shows that an element $t \in S$ satisfies $f(t) = 1$ **if and only if** $t \notin X$. Thus, for each $t \in S$, we have the following chain of equivalences:

$$(f(t) = 1) \iff (t \notin X) \iff (t \in S \setminus X) \iff (t \in Y)$$

(since $S \setminus X = Y$). In other words, an element $t \in S$ satisfies $f(t) = 1$ **if and only if** t belongs to Y. Hence, the set of all elements $t \in S$ that satisfy $f(t) = 1$ is precisely Y. In other words,

$$\{t \in S \mid f(t) = 1\} = Y.$$

But the definition of b_1 yields

$$b_1 = \sum_{\substack{s \in S; \\ f(s)=1}} a_s = \sum_{s \in \{t \in S \mid f(t)=1\}} a_s = \sum_{s \in Y} a_s$$

(since $\{t \in S \mid f(t) = 1\} = Y$). This proves (200).]

Now, (198) becomes

$$\sum_{s \in S} a_s = \underbrace{b_0}_{\substack{= \sum_{s \in X} a_s \\ \text{(by (199))}}} + \underbrace{b_1}_{\substack{= \sum_{s \in Y} a_s \\ \text{(by (200))}}} = \sum_{s \in X} a_s + \sum_{s \in Y} a_s.$$

This proves Theorem 2.130. □

Similarly, we can prove the equality (26). (This proof was already outlined in Section 1.4.)

A consequence of Theorem 2.130 is the following fact, which has appeared as the equality (11) in Section 1.4:

Corollary 2.131. Let S be a finite set. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Let T be a subset of S such that every $s \in T$ satisfies $a_s = 0$. Then,

$$\sum_{s \in S} a_s = \sum_{s \in S \setminus T} a_s.$$

Proof of Corollary 2.131. We have assumed that every $s \in T$ satisfies $a_s = 0$. Thus, $\sum_{s \in T} \underbrace{a_s}_{=0} = \sum_{s \in T} 0 = 0$ (by Theorem 2.126 (applied to T instead of S)).

But T and $S \setminus T$ are subsets of S. These two subsets satisfy $T \cap (S \setminus T) = \varnothing$ and $T \cup (S \setminus T) = S$ (since $T \subseteq S$). Hence, Theorem 2.130 (applied to $X = T$ and $Y = S \setminus T$) yields

$$\sum_{s \in S} a_s = \underbrace{\sum_{s \in T} a_s}_{=0} + \sum_{s \in S \setminus T} a_s = \sum_{s \in S \setminus T} a_s.$$

This proves Corollary 2.131. $\square$

2.14.12. Substituting the summation index

Next, we shall show the equality (12):

Theorem 2.132. Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_t be an element of $\mathbb{A}$ for each $t \in T$. Then,

$$\sum_{t \in T} a_t = \sum_{s \in S} a_{f(s)}.$$

Proof of Theorem 2.132. Each $w \in T$ satisfies

$$\sum_{\substack{s \in S; \\ f(s) = w}} a_{f(s)} = a_w. \tag{201}$$

[*Proof of (201):* Let $w \in T$.

The map f is bijective; thus, it is invertible. In other words, its inverse map $f^{-1} : T \to S$ exists. Hence, $f^{-1}(w)$ is a well-defined element of S, and is the only element $t \in S$ satisfying $f(t) = w$. Therefore,

$$\{t \in S \mid f(t) = w\} = \left\{ f^{-1}(w) \right\}. \tag{202}$$

Now,

$$\begin{aligned}
&\sum_{\substack{s \in S; \\ f(s) = w}} a_{f(s)} \\
&= \sum_{s \in \{t \in S \mid f(t) = w\}} a_{f(s)} = \sum_{s \in \left\{ f^{-1}(w) \right\}} a_{f(s)} \qquad \text{(by (202))} \\
&= a_{f\left(f^{-1}(w)\right)} \qquad \left(\begin{array}{c} \text{by Proposition 2.121 \textbf{(a)} (applied to } \left\{ f^{-1}(w) \right\}, a_{f(s)} \\ \text{and } f^{-1}(w) \text{ instead of } S, a_s \text{ and } p) \end{array} \right) \\
&= a_w \qquad \left(\text{since } f\left(f^{-1}(w) \right) = w \right).
\end{aligned}$$

This proves (201).]

Renaming the summation index w as t in the sum $\sum_{w \in T} a_w$ does not change the sum (since $(a_w)_{w \in T}$ and $(a_t)_{t \in T}$ are the same $\mathbb{A}$-valued T-family). In other words, $\sum_{w \in T} a_w = \sum_{t \in T} a_t$.

Theorem 2.127 (applied to T and $a_{f(s)}$ instead of W and a_s) yields

$$\sum_{s \in S} a_{f(s)} = \sum_{w \in T} \underbrace{\sum_{\substack{s \in S; \\ f(s) = w}} a_{f(s)}}_{\substack{= a_w \\ \text{(by (201))}}} = \sum_{w \in T} a_w = \sum_{t \in T} a_t.$$

This proves Theorem 2.132. □

2.14.13. Sums of congruences

Proposition 2.21 **(a)** says that we can add two congruences modulo an integer n. We shall now see that we can add **any** number of congruences modulo an integer n:

Theorem 2.133. Let n be an integer. Let S be a finite set. For each $s \in S$, let a_s and b_s be two integers. Assume that

$$a_s \equiv b_s \mod n \qquad \text{for each } s \in S.$$

Then,

$$\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n.$$

Proof of Theorem 2.133. We forget that we fixed n, S, a_s and b_s. We shall prove Theorem 2.133 by induction on $|S|$:

Induction base: The induction base (i.e., proving that Theorem 2.133 holds under the condition that $|S| = 0$) is left to the reader (as it boils down to the trivial fact that $0 \equiv 0 \mod n$).

Induction step: Let $m \in \mathbb{N}$. Assume that Theorem 2.133 holds under the condition that $|S| = m$. We must now prove that Theorem 2.133 holds under the condition that $|S| = m + 1$.

We have assumed that Theorem 2.133 holds under the condition that $|S| = m$. In other words, the following claim holds:

> *Claim 1:* Let n be an integer. Let S be a finite set such that $|S| = m$. For each $s \in S$, let a_s and b_s be two integers. Assume that
>
> $$a_s \equiv b_s \mod n \qquad \text{for each } s \in S.$$
>
> Then,
>
> $$\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n.$$

Next, we shall show the following claim:

Claim 2: Let n be an integer. Let S be a finite set such that $|S| = m+1$. For each $s \in S$, let a_s and b_s be two integers. Assume that

$$a_s \equiv b_s \mod n \qquad \text{for each } s \in S. \tag{203}$$

Then,

$$\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n.$$

[*Proof of Claim 2:* We have $|S| = m+1 > m \geq 0$. Hence, the set S is nonempty. Thus, there exists some $t \in S$. Consider this t.

From $t \in S$, we obtain $|S \setminus \{t\}| = |S| - 1 = m$ (since $|S| = m+1$). Also, every $s \in S \setminus \{t\}$ satisfies $s \in S \setminus \{t\} \subseteq S$ and thus $a_s \equiv b_s \mod n$ (by (203)). In other words, we have

$$a_s \equiv b_s \mod n \qquad \text{for each } s \in S \setminus \{t\}.$$

Hence, Claim 1 (applied to $S \setminus \{t\}$ instead of S) yields

$$\sum_{s \in S \setminus \{t\}} a_s \equiv \sum_{s \in S \setminus \{t\}} b_s \mod n. \tag{204}$$

But $t \in S$. Hence, (203) (applied to $s = t$) yields $a_t \equiv b_t \mod n$.

Now, Proposition 2.120 (applied to b_s instead of a_s) yields

$$\sum_{s \in S} b_s = b_t + \sum_{s \in S \setminus \{t\}} b_s. \tag{205}$$

But Proposition 2.120 yields

$$\sum_{s \in S} a_s = \underbrace{a_t}_{\equiv b_t \mod n} + \underbrace{\sum_{s \in S \setminus \{t\}} a_s}_{\substack{\equiv \sum_{s \in S \setminus \{t\}} b_s \mod n \\ \text{(by (204))}}} \equiv b_t + \sum_{s \in S \setminus \{t\}} b_s = \sum_{s \in S} b_s \mod n$$

(by (205)). This proves Claim 2.]

But Claim 2 says precisely that Theorem 2.133 holds under the condition that $|S| = m+1$. Hence, we conclude that Theorem 2.133 holds under the condition that $|S| = m+1$ (since Claim 2 is proven). This completes the induction step. Thus, Theorem 2.133 is proven by induction. □

As we said, Theorem 2.133 shows that we can sum up several congruences. Thus, we can extend our principle of substitutivity for congruences as follows:

> *Principle of substitutivity for congruences (stronger version):* Fix an integer n. If two numbers x and x' are congruent to each other modulo n (that is, $x \equiv x' \mod n$), and if we have any expression A that involves only integers, addition, subtraction, multiplication **and summation signs**, and involves the object x, then we can replace this x (or, more precisely, any arbitrary appearance of x in A) in A by x'; the value of the resulting expression A' will be congruent to the value of A modulo n.

For example, if $p \in \mathbb{N}$, then

$$\sum_{s \in \{1,2,\ldots,p\}} s^2 (5 - 3s) \equiv \sum_{s \in \{1,2,\ldots,p\}} s (5 - 3s) \mod 2$$

(here, we have replaced the "s^2" inside the sum by "s"), because every $s \in \{1, 2, \ldots, p\}$ satisfies $s^2 \equiv s \mod 2$ (this is easy to check[85]).

2.14.14. Finite products

Proposition 2.103 is a property of the addition of numbers; it has an analogue for multiplication of numbers:

Proposition 2.134. Let a, b and c be three numbers (i.e., elements of $\mathbb{A}$). Then, $(ab) c = a (bc)$.

Proposition 2.134 is known as the *associativity of multiplication* (in $\mathbb{A}$), and is fundamental; its proof can be found in any textbook on the construction of the number system[86].

Proposition 2.104 also has an analogue for multiplication:

Proposition 2.135. Let a and b be two numbers (i.e., elements of $\mathbb{A}$). Then, $ab = ba$.

Proposition 2.135 is known as the *commutativity of multiplication* (in $\mathbb{A}$), and again is a fundamental result whose proofs are found in standard textbooks[87].

---

[85] *Proof.* Let $p \in \mathbb{N}$ and $s \in \{1, 2, \ldots, p\}$. We must prove that $s^2 \equiv s \mod 2$.

We have $s \in \{1, 2, \ldots, p\}$ and thus $s - 1 \in \{0, 1, \ldots, p - 1\} \subseteq \mathbb{N}$. Hence, (172) (applied to $n = s - 1$) yields $\sum_{i \in \{1,2,\ldots,s-1\}} i = \frac{(s-1)((s-1)+1)}{2} = \frac{(s-1)s}{2}$. Hence, $\frac{(s-1)s}{2}$ is an integer (since $\sum_{i \in \{1,2,\ldots,s-1\}} i$ is an integer). In other words, $2 \mid (s-1)s$. In other words, $2 \mid s^2 - s$ (since $(s-1)s = s^2 - s$). In other words, $s^2 \equiv s \mod 2$ (by the definition of "congruent"), qed.

[86] For example, Proposition 2.134 is proven in [Swanso18, Theorem 3.2.3 (7)] for the case when $\mathbb{A} = \mathbb{N}$; in [Swanso18, Theorem 3.5.4 (7)] for the case when $\mathbb{A} = \mathbb{Z}$; in [Swanso18, Theorem 3.6.4 (7)] for the case when $\mathbb{A} = \mathbb{Q}$; in [Swanso18, Theorem 3.7.13] for the case when $\mathbb{A} = \mathbb{R}$; in [Swanso18, Theorem 3.9.3] for the case when $\mathbb{A} = \mathbb{C}$.

[87] For example, Proposition 2.135 is proven in [Swanso18, Theorem 3.2.3 (8)] for the case when $\mathbb{A} = \mathbb{N}$; in [Swanso18, Theorem 3.5.4 (8)] for the case when $\mathbb{A} = \mathbb{Z}$; in [Swanso18, Theorem 3.6.4 (8)] for the case when $\mathbb{A} = \mathbb{Q}$; in [Swanso18, Theorem 3.7.13] for the case when $\mathbb{A} = \mathbb{R}$; in [Swanso18, Theorem 3.9.3] for the case when $\mathbb{A} = \mathbb{C}$.

Proposition 2.125 has an analogue for multiplication as well (but note that x now needs to be in $\mathbb{N}$, in order to guarantee that the powers are well-defined):

Proposition 2.136. Let $x \in \mathbb{N}$. Let y and z be two numbers (i.e., elements of $\mathbb{A}$). Then, $(yz)^x = y^x z^x$.

Proposition 2.136 is one of the laws of exponents, and can easily be shown by induction on x (using Proposition 2.135 and Proposition 2.134).

So far in Section 2.14, we have been studying **sums** of $\mathbb{A}$-valued S-families (when S is a finite set): We have proven that the definition of $\sum_{s \in S} a_s$ given in Section 1.4 is legitimate, and we have proven several properties of such sums. By the exact same reasoning (but with addition replaced by multiplication), we can study **products** of $\mathbb{A}$-valued S-families. In particular, we can similarly prove that the definition of $\prod_{s \in S} a_s$ given in Section 1.4 is legitimate, and we can prove properties of such products that are analogous to the properties of sums proven above (except for Proposition 2.119, which does not have an analogue for products)[88]. For example, the following theorems are analogues of Theorem 2.122, Theorem 2.124, Theorem 2.126, Theorem 2.127, Theorem 2.132 and Theorem 2.133, respectively:

Theorem 2.137. Let S be a finite set. For every $s \in S$, let a_s and b_s be elements of $\mathbb{A}$. Then,

$$\prod_{s \in S} (a_s b_s) = \left(\prod_{s \in S} a_s \right) \cdot \left(\prod_{s \in S} b_s \right).$$

[88] We need to be slightly careful when we adapt our above proofs to products instead of sums: Apart from replacing addition by multiplication everywhere, we need to:

- replace the number 0 by 1 whenever it appears in a computation inside $\mathbb{A}$ (but, of course, not when it appears as the size of a set);
- replace every $\sum$ sign by a $\prod$ sign;
- replace "let λ be an element of $\mathbb{A}$" by "let λ be an element of $\mathbb{N}$" in Theorem 2.124;
- replace any expression of the form "λb" by "b^λ" in Theorem 2.124 (so that the claim of Theorem 2.124 becomes $\prod_{s \in S} (a_s)^\lambda = \left(\prod_{s \in S} a_s \right)^\lambda$) and in its proof;
- replace every reference to Proposition 2.103 by a reference to Proposition 2.134;
- replace every reference to Proposition 2.104 by a reference to Proposition 2.135;
- replace every reference to Proposition 2.125 by a reference to Proposition 2.136.

And, to be fully precise: We should not replace addition by multiplication **everywhere** (e.g., we should not replace "$|S| = m + 1$" by "$|S| = m \cdot 1$" in the proof of Theorem 2.127), but of course only where it stands for the addition **inside** $\mathbb{A}$.

Theorem 2.138. Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{A}$. Also, let λ be an element of $\mathbb{N}$. Then,

$$\prod_{s\in S} (a_s)^\lambda = \left(\prod_{s\in S} a_s \right)^\lambda .$$

Theorem 2.139. Let S be a finite set. Then,

$$\prod_{s\in S} 1 = 1.$$

Theorem 2.140. Let S be a finite set. Let W be a finite set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then,

$$\prod_{s\in S} a_s = \prod_{w\in W} \prod_{\substack{s\in S;\\ f(s)=w}} a_s.$$

Theorem 2.141. Let S and T be two finite sets. Let $f : S \to T$ be a **bijective** map. Let a_t be an element of $\mathbb{A}$ for each $t \in T$. Then,

$$\prod_{t\in T} a_t = \prod_{s\in S} a_{f(s)}.$$

Theorem 2.142. Let n be an integer. Let S be a finite set. For each $s \in S$, let a_s and b_s be two integers. Assume that

$$a_s \equiv b_s \mod n \qquad \text{for each } s \in S.$$

Then,

$$\prod_{s\in S} a_s \equiv \prod_{s\in S} b_s \mod n.$$

2.14.15. Finitely supported (but possibly infinite) sums

In Section 1.4, we mentioned that a sum of the form $\sum_{s\in S} a_s$ can be well-defined even when the set S is not finite. Indeed, for it to be well-defined, it suffices that **only finitely many among the a_s are nonzero** (or, more rigorously: only finitely many $s \in S$ satisfy $a_s \neq 0$). As we already mentioned, the sum $\sum_{s\in S} a_s$ in this case is defined by discarding the zero addends and summing the finitely many addends that remain. Let us briefly discuss such sums (without focussing on advanced properties):

Definition 2.143. Let S be any set. An $\mathbb{A}$-valued S-family $(a_s)_{s\in S}$ is said to be *finitely supported* if only finitely many $s \in S$ satisfy $a_s \neq 0$.

So the sums we want to discuss are sums $\sum\limits_{s\in S} a_s$ for which the set S may be infinite but the S-family $(a_s)_{s\in S}$ is finitely supported. Let us repeat the definition of such sums in more rigorous language:

Definition 2.144. Let S be any set. Let $(a_s)_{s\in S}$ be a finitely supported $\mathbb{A}$-valued S-family. Thus, there exists a **finite** subset T of S such that

$$\text{every } s \in S \setminus T \text{ satisfies } a_s = 0. \tag{206}$$

(This is because only finitely many $s \in S$ satisfy $a_s \neq 0$.) We then define the sum $\sum\limits_{s\in S} a_s$ to be $\sum\limits_{s\in T} a_s$. (This definition is legitimate, because Proposition 2.145 **(a)** below shows that $\sum\limits_{s\in T} a_s$ does not depend on the choice of T.)

This definition formalizes what we said above about making sense of $\sum\limits_{s\in S} a_s$: Namely, we discard zero addends (namely, the addends corresponding to $s \in S \setminus T$) and only sum the finitely many addends that remain (these are the addends corresponding to $s \in T$); thus, we get $\sum\limits_{s\in T} a_s$. Note that we are not requiring that every $s \in T$ satisfies $a_s \neq 0$; that is, we are not necessarily discarding **all** the zero addends from our sum (but merely discarding enough of them to ensure that only finitely many remain). This may appear like a strange choice (why introduce extra freedom into the definition?), but is reasonable from the viewpoint of constructive mathematics (where it is not always decidable if a number is 0 or not).

Proposition 2.145. Let S be any set. Let $(a_s)_{s\in S}$ be a finitely supported $\mathbb{A}$-valued S-family.

(a) If T is a finite subset of S such that (206) holds, then the sum $\sum\limits_{s\in T} a_s$ does not depend on the choice of T. (That is, if T_1 and T_2 are two finite subsets T of S satisfying (206), then $\sum\limits_{s\in T_1} a_s = \sum\limits_{s\in T_2} a_s$.)

(b) If the set S is finite, then the sum $\sum\limits_{s\in S} a_s$ defined in Definition 2.144 is identical with the sum $\sum\limits_{s\in S} a_s$ defined in Definition 2.111. (Thus, Definition 2.144 does not conflict with the previous definition of $\sum\limits_{s\in S} a_s$ for finite sets S.)

Proposition 2.145 is fairly easy to prove using Corollary 2.131; this proof is part of Exercise 2.3 below.

Most properties of finite sums have analogues for sums of finitely supported $\mathbb{A}$-valued S-families. For example, here is an analogue of Theorem 2.122:

Theorem 2.146. Let S be a set. Let $(a_s)_{s\in S}$ and $(b_s)_{s\in S}$ be two finitely supported $\mathbb{A}$-valued S-families. Then, the $\mathbb{A}$-valued S-family $(a_s + b_s)_{s\in S}$ is finitely supported as well, and we have

$$\sum_{s\in S} (a_s + b_s) = \sum_{s\in S} a_s + \sum_{s\in S} b_s.$$

The proof of Theorem 2.146 is fairly simple (it relies prominently on the fact that the union of two finite sets is finite), and again is part of Exercise 2.3 below.

It is also easy to state and prove analogues of Theorem 2.124 and Theorem 2.126.

We can next prove (22) in full generality (not only when W is finite):

Theorem 2.147. Let S be a finite set. Let W be a set. Let $f : S \to W$ be a map. Let a_s be an element of $\mathbb{A}$ for each $s \in S$. Then, the $\mathbb{A}$-valued W-family $\left(\sum_{\substack{s\in S;\\ f(s)=w}} a_s \right)_{w\in W}$ is finitely supported and satisfies

$$\sum_{s\in S} a_s = \sum_{w\in W} \sum_{\substack{s\in S;\\ f(s)=w}} a_s.$$

Note that the sum on the right hand side of Theorem 2.147 makes sense even when W is infinite, because the W-family $\left(\sum_{\substack{s\in S;\\ f(s)=w}} a_s \right)_{w\in W}$ is finitely supported (i.e., only finitely many $w \in W$ satisfy $\sum_{\substack{s\in S;\\ f(s)=w}} a_s \neq 0$). The easiest way to prove Theorem 2.147 is probably by reducing it to Theorem 2.127 (since $f(S)$ is a finite subset of W, and every $w \in W \setminus f(S)$ satisfies $\sum_{\substack{s\in S;\\ f(s)=w}} a_s = (\text{empty sum}) = 0$). Again, we leave the details to the interested reader.

Again, we refer to Exercise 2.3 for the proof of Theorem 2.147.

Actually, Theorem 2.147 can be generalized even further:

Theorem 2.148. Let S be a set. Let W be a set. Let $f : S \to W$ be a map. Let $(a_s)_{s\in S}$ be a finitely supported $\mathbb{A}$-valued S-family. Then, for each $w \in W$, the $\mathbb{A}$-valued $\{t \in S \mid f(t) = w\}$-family $(a_s)_{s\in\{t\in S \mid f(t)=w\}}$ is finitely supported as well (so that the sum $\sum\limits_{\substack{s\in S;\\ f(s)=w}} a_s$ is well-defined). Furthermore, the $\mathbb{A}$-valued W-family $\left(\sum\limits_{\substack{s\in S;\\ f(s)=w}} a_s\right)_{w\in W}$ is also finitely supported. Finally,

$$\sum_{s\in S} a_s = \sum_{w\in W} \sum_{\substack{s\in S;\\ f(s)=w}} a_s.$$

Again, see Exercise 2.3 for the proof. This theorem can be used to obtain an analogue of Theorem 2.130 for finitely supported $\mathbb{A}$-valued S-families.

Exercise 2.3. Prove Proposition 2.145, Theorem 2.146, Theorem 2.147 and Theorem 2.148.

Thus, we have defined the values of certain infinite sums (although not nearly as many infinite sums as analysis can make sense of). We can similarly define the values of certain infinite products: In order for $\prod\limits_{s\in S} a_s$ to be well-defined, it suffices that **only finitely many among the** a_s **are distinct from** 1 (or, more rigorously: only finitely many $s \in S$ satisfy $a_s \neq 1$). We leave the details and properties of this definition to the reader.

2.15. Two-sided induction

2.15.1. The principle of two-sided induction

Let us now return to studying induction principles. We have seen several induction principles that allow us to prove statements about nonnegative integers, integers in $\mathbb{Z}_{\geq g}$ or integers in an interval. What about proving statements about **arbitrary** integers? The induction principles we have seen so far do not suffice to prove such statements directly, since our induction steps always "go up" (in the sense that they begin by assuming that our statement $\mathcal{A}(k)$ holds for some integers k, and involve proving that it also holds for a **larger** value of k), but it is impossible to traverse all the integers by starting at some integer g and going up (you will never get to $g-1$ this way). In contrast, the following induction principle includes both an "upwards" and a "downwards" induction step, which makes it suited for proving statements about all integers:

Theorem 2.149. Let $g \in \mathbb{Z}$. Let $\mathbb{Z}_{\leq g}$ be the set $\{g, g-1, g-2, \ldots\}$ (that is, the set of all integers that are $\leq g$).

For each $n \in \mathbb{Z}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: The statement $\mathcal{A}(g)$ holds.

Assumption 2: If $m \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds.

Assumption 3: If $m \in \mathbb{Z}_{\leq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m-1)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}$.

Theorem 2.149 is known as the *principle of two-sided induction*. Roughly speaking, a proof using Theorem 2.149 will involve two induction steps: one that "goes up" (proving that Assumption 2 holds), and one that "goes down" (proving that Assumption 3 holds). However, in practice, Theorem 2.149 is seldom used, which is why we shall not make any conventions about how to write proofs using Theorem 2.149. We will only give one example for such a proof.

Let us first prove Theorem 2.149 itself:

Proof of Theorem 2.149. Assumptions 1 and 2 of Theorem 2.149 are exactly Assumptions 1 and 2 of Theorem 2.53. Hence, Assumptions 1 and 2 of Theorem 2.53 hold (since Assumptions 1 and 2 of Theorem 2.149 hold). Thus, Theorem 2.53 shows that

$$\mathcal{A}(n) \text{ holds for each } n \in \mathbb{Z}_{\geq g}. \tag{207}$$

On the other hand, for each $n \in \mathbb{Z}$, we define a logical statement $\mathcal{B}(n)$ by $\mathcal{B}(n) = \mathcal{A}(2g-n)$. We shall now consider the Assumptions A and B of Corollary 2.61.

The definition of $\mathcal{B}(g)$ yields $\mathcal{B}(g) = \mathcal{A}(2g-g) = \mathcal{A}(g)$ (since $2g - g = g$). Hence, the statement $\mathcal{B}(g)$ holds (since the statement $\mathcal{A}(g)$ holds (by Assumption 1)). In other words, Assumption A is satisfied.

Next, let $p \in \mathbb{Z}_{\geq g}$ be such that $\mathcal{B}(p)$ holds. We shall show that $\mathcal{B}(p+1)$ holds.

Indeed, we have $\mathcal{B}(p) = \mathcal{A}(2g-p)$ (by the definition of $\mathcal{B}(p)$). Thus, $\mathcal{A}(2g-p)$ holds (since $\mathcal{B}(p)$ holds). But $p \in \mathbb{Z}_{\geq g}$; hence, p is an integer that is $\geq g$. Thus, $p \geq g$, so that $2g - \underbrace{p}_{\geq g} \leq 2g - g = g$. Hence, $2g-p$ is an integer that is $\leq g$. In other words, $2g - p \in \mathbb{Z}_{\leq g}$. Therefore, Assumption 3 (applied to $m = 2g-p$) shows that $\mathcal{A}(2g-p-1)$ also holds (since $\mathcal{A}(2g-p)$ holds). But the definition of $\mathcal{B}(p+1)$ yields $\mathcal{B}(p+1) = \mathcal{A}\left(\underbrace{2g-(p+1)}_{=2g-p-1}\right) = \mathcal{A}(2g-p-1)$. Hence, $\mathcal{B}(p+1)$ holds (since $\mathcal{A}(2g-p-1)$ holds).

Now, forget that we fixed p. We thus have shown that if $p \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds. In other words, Assumption B is satisfied.

We now have shown that both Assumptions A and B are satisfied. Hence, Corollary 2.61 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{Z}_{\geq g}. \tag{208}$$

Now, let $n \in \mathbb{Z}$. We shall prove that $\mathcal{A}(n)$ holds.

Indeed, we have either $n \geq g$ or $n < g$. Hence, we are in one of the following two cases:

Case 1: We have $n \geq g$.

Case 2: We have $n < g$.

Let us first consider Case 1. In this case, we have $n \geq g$. Hence, $n \in \mathbb{Z}_{\geq g}$ (since n is an integer). Thus, (207) shows that $\mathcal{A}(n)$ holds. We thus have proven that $\mathcal{A}(n)$ holds in Case 1.

Let us now consider Case 2. In this case, we have $n < g$. Thus, $n \leq g$. Hence, $2g - \underbrace{n}_{\leq g} \geq 2g - g = g$. Thus, $2g - n \in \mathbb{Z}_{\geq g}$ (since $2g - n$ is an integer). Hence, (208) (applied to $2g - n$ instead of n) shows that $\mathcal{B}(2g - n)$ holds. But the definition of $\mathcal{B}(2g - n)$ yields $\mathcal{B}(2g - n) = \mathcal{A}\left(\underbrace{2g - (2g - n)}_{=n}\right) = \mathcal{A}(n)$. Hence, $\mathcal{A}(n)$ holds (since $\mathcal{B}(2g - n)$ holds). Thus, we have proven that $\mathcal{A}(n)$ holds in Case 2.

We now have shown that $\mathcal{A}(n)$ holds in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that $\mathcal{A}(n)$ always holds.

Now, forget that we fixed n. We thus have proven that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}$. This proves Theorem 2.149. □

As an example for the use of Theorem 2.149, we shall prove the following fact:

Proposition 2.150. Let N be a positive integer. For each $n \in \mathbb{Z}$, there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $n = qN + r$.

We shall soon see (in Theorem 2.153) that these q and r are actually uniquely determined by N and n; they are called the *quotient* and the *remainder of the division of n by N*. This is fundamental to all of number theory.

Example 2.151. If we apply Proposition 2.150 to $N = 4$ and $n = 10$, then we conclude that there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, 2, 3\}$ such that $10 = q \cdot 4 + r$. And indeed, such q and r can easily be found ($q = 2$ and $r = 2$).

Proof of Proposition 2.150. First, we notice that $N - 1 \in \mathbb{N}$ (since N is a positive integer). Hence, $0 \in \{0, 1, \ldots, N-1\}$ and $N - 1 \in \{0, 1, \ldots, N-1\}$.

For each $n \in \mathbb{Z}$, we let $\mathcal{A}(n)$ be the statement

$$(\text{there exist } q \in \mathbb{Z} \text{ and } r \in \{0, 1, \ldots, N-1\} \text{ such that } n = qN + r).$$

Let $g = 0$; thus, $g \in \mathbb{Z}$. Define the set $\mathbb{Z}_{\leq g}$ as in Theorem 2.149. (Thus, $\mathbb{Z}_{\leq g} = \{g, g-1, g-2, \ldots\} = \{0, -1, -2, \ldots\}$ is the set of all nonpositive integers.) We shall now show that Assumptions 1, 2 and 3 of Theorem 2.149 are satisfied.

[*Proof that Assumption 1 is satisfied:* We have $0 \in \{0, 1, \ldots, N-1\}$ and $0 = 0N + 0$. Hence, there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $0 = qN + r$ (namely,

$q = 0$ and $r = 0$). In other words, the statement $\mathcal{A}(0)$ holds[89]. In other words, the statement $\mathcal{A}(g)$ holds (since $g = 0$). In other words, Assumption 1 is satisfied.]

[*Proof that Assumption 2 is satisfied:* Let $m \in \mathbb{Z}_{\geq g}$ be such that $\mathcal{A}(m)$ holds. We shall show that $\mathcal{A}(m+1)$ also holds.

We know that $\mathcal{A}(m)$ holds. In other words, there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m = qN + r$ [90]. Consider these q and r, and denote them by q_0 and r_0. Thus, $q_0 \in \mathbb{Z}$ and $r_0 \in \{0, 1, \ldots, N-1\}$ and $m = q_0 N + r_0$.

Now, we are in one of the following two cases:

Case 1: We have $r_0 = N - 1$.

Case 2: We have $r_0 \neq N - 1$.

Let us first consider Case 1. In this case, we have $r_0 = N - 1$. Hence, $r_0 + 1 = N$. Now,

$$\underbrace{m}_{= q_0 N + r_0} + 1 = q_0 N + \underbrace{r_0 + 1}_{= N} = q_0 N + N = (q_0 + 1) N = (q_0 + 1) N + 0.$$

Since $0 \in \{0, 1, \ldots, N-1\}$, we can therefore conclude that there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m + 1 = qN + r$ (namely, $q = q_0 + 1$ and $r = 0$). In other words, the statement $\mathcal{A}(m+1)$ holds[91]. Thus, we have proven that $\mathcal{A}(m+1)$ holds in Case 1.

Let us next consider Case 2. In this case, we have $r_0 \neq N - 1$. Combining $r_0 \in \{0, 1, \ldots, N-1\}$ with $r_0 \neq N - 1$, we obtain

$$r_0 \in \{0, 1, \ldots, N-1\} \setminus \{N-1\} = \{0, 1, \ldots, N-2\},$$

so that $r_0 + 1 \in \{1, 2, \ldots, N-1\} \subseteq \{0, 1, \ldots, N-1\}$. Also,

$$\underbrace{m}_{= q_0 N + r_0} + 1 = q_0 N + r_0 + 1 = q_0 N + (r_0 + 1).$$

Since $r_0 + 1 \in \{0, 1, \ldots, N-1\}$, we can therefore conclude that there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m + 1 = qN + r$ (namely, $q = q_0$ and $r = r_0 + 1$). In other words, the statement $\mathcal{A}(m+1)$ holds[92]. Thus, we have proven that $\mathcal{A}(m+1)$ holds in Case 2.

We have now proven that $\mathcal{A}(m+1)$ holds in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that $\mathcal{A}(m+1)$ always holds.

---

[89]since the statement $\mathcal{A}(0)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $0 = qN + r$)

[90]since the statement $\mathcal{A}(m)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m = qN + r$)

[91]since the statement $\mathcal{A}(m+1)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m + 1 = qN + r$)

[92]since the statement $\mathcal{A}(m+1)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m + 1 = qN + r$)

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m+1)$ also holds. In other words, Assumption 2 is satisfied.]

[*Proof that Assumption 3 is satisfied:* Let $m \in \mathbb{Z}_{\leq g}$ be such that $\mathcal{A}(m)$ holds. We shall show that $\mathcal{A}(m-1)$ also holds.

We know that $\mathcal{A}(m)$ holds. In other words, there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m = qN + r$ [93]. Consider these q and r, and denote them by q_0 and r_0. Thus, $q_0 \in \mathbb{Z}$ and $r_0 \in \{0, 1, \ldots, N-1\}$ and $m = q_0 N + r_0$.

Now, we are in one of the following two cases:

Case 1: We have $r_0 = 0$.

Case 2: We have $r_0 \neq 0$.

Let us first consider Case 1. In this case, we have $r_0 = 0$. Now,

$$\underbrace{m}_{=q_0 N + r_0} - 1 = q_0 N + \underbrace{r_0}_{=0} - 1 = q_0 N - 1 = \underbrace{q_0 N - N}_{=(q_0 - 1)N} + (N-1) = (q_0 - 1) N + (N-1).$$

Since $N - 1 \in \{0, 1, \ldots, N-1\}$, we can therefore conclude that there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m - 1 = qN + r$ (namely, $q = q_0 - 1$ and $r = N - 1$). In other words, the statement $\mathcal{A}(m-1)$ holds[94]. Thus, we have proven that $\mathcal{A}(m-1)$ holds in Case 1.

Let us next consider Case 2. In this case, we have $r_0 \neq 0$. Combining $r_0 \in \{0, 1, \ldots, N-1\}$ with $r_0 \neq 0$, we obtain

$$r_0 \in \{0, 1, \ldots, N-1\} \setminus \{0\} = \{1, 2, \ldots, N-1\},$$

so that $r_0 - 1 \in \{0, 1, \ldots, N-2\} \subseteq \{0, 1, \ldots, N-1\}$. Also,

$$\underbrace{m}_{=q_0 N + r_0} - 1 = q_0 N + r_0 - 1 = q_0 N + (r_0 - 1).$$

Since $r_0 - 1 \in \{0, 1, \ldots, N-1\}$, we can therefore conclude that there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m - 1 = qN + r$ (namely, $q = q_0$ and $r = r_0 - 1$). In other words, the statement $\mathcal{A}(m-1)$ holds[95]. Thus, we have proven that $\mathcal{A}(m-1)$ holds in Case 2.

We have now proven that $\mathcal{A}(m-1)$ holds in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that $\mathcal{A}(m-1)$ always holds.

Now, forget that we fixed m. We thus have shown that if $m \in \mathbb{Z}_{\leq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m-1)$ also holds. In other words, Assumption 3 is satisfied.]

[93] since the statement $\mathcal{A}(m)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m = qN + r$)

[94] since the statement $\mathcal{A}(m-1)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m - 1 = qN + r$)

[95] since the statement $\mathcal{A}(m-1)$ is defined as
(there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $m - 1 = qN + r$)

We have now shown that all three Assumptions 1, 2 and 3 of Theorem 2.149 are satisfied. Thus, Theorem 2.149 yields that

$$\mathcal{A}(n) \text{ holds for each } n \in \mathbb{Z}. \tag{209}$$

Now, let $n \in \mathbb{Z}$. Then, $\mathcal{A}(n)$ holds (by (209)). In other words, there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $n = qN + r$ (because the statement $\mathcal{A}(n)$ is defined as (there exist $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, N-1\}$ such that $n = qN + r$)). This proves Proposition 2.150. □

2.15.2. Division with remainder

We need one more lemma:

Lemma 2.152. Let N be a positive integer. Let $(q_1, r_1) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$ and $(q_2, r_2) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$. Assume that $q_1 N + r_1 = q_2 N + r_2$. Then, $(q_1, r_1) = (q_2, r_2)$.

Proof of Lemma 2.152. We have $(q_1, r_1) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$. Thus, $q_1 \in \mathbb{Z}$ and $r_1 \in \{0, 1, \ldots, N-1\}$. Similarly, $q_2 \in \mathbb{Z}$ and $r_2 \in \{0, 1, \ldots, N-1\}$.

From $r_1 \in \{0, 1, \ldots, N-1\}$, we obtain $r_1 \geq 0$. From $r_2 \in \{0, 1, \ldots, N-1\}$, we obtain $r_2 \leq N-1$. Hence, $\underbrace{r_2}_{\leq N-1} - \underbrace{r_1}_{\geq 0} \leq (N-1) - 0 = N - 1 < N$.

Next, we shall prove that

$$q_1 \leq q_2. \tag{210}$$

[*Proof of (210):* Assume the contrary. Thus, $q_1 > q_2$, so that $q_1 - q_2 > 0$. Hence, $q_1 - q_2 \geq 1$ (since $q_1 - q_2$ is an integer). Therefore, $q_1 - q_2 - 1 \geq 0$, so that $N(q_1 - q_2 - 1) \geq 0$ (because $N > 0$ and $q_1 - q_2 - 1 \geq 0$).

But $q_1 N + r_1 = q_2 N + r_2$, so that $q_2 N + r_2 = q_1 N + r_1$. Hence,

$$r_2 - r_1 = q_1 N - q_2 N = N(q_1 - q_2) = \underbrace{N(q_1 - q_2 - 1)}_{\geq 0} + N \geq N.$$

This contradicts $r_2 - r_1 < N$. This contradiction shows that our assumption was wrong. Hence, (210) is proven.]

Thus, we have proven that $q_1 \leq q_2$. The same argument (with the roles of (q_1, r_1) and (q_2, r_2) interchanged) shows that $q_2 \leq q_1$. Combining the inequalities $q_1 \leq q_2$ and $q_2 \leq q_1$, we obtain $q_1 = q_2$.

Also, $q_2 N + r_2 = \underbrace{q_1}_{= q_2} N + r_1 = q_2 N + r_1$. Subtracting $q_2 N$ from both sides of this equality, we find $r_2 = r_1$. Hence, $r_1 = r_2$.

Thus, $\left(\underbrace{q_1}_{= q_2}, \underbrace{r_1}_{= r_2}\right) = (q_2, r_2)$. This proves Lemma 2.152. □

As we have already mentioned, Proposition 2.150 is just a part of a crucial result from number theory:

Theorem 2.153. Let N be a positive integer. Let $n \in \mathbb{Z}$. Then, there is a unique pair $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$.

Proving this theorem will turn out rather easy, since we have already done the hard work with our proofs of Proposition 2.150 and Lemma 2.152:

Proof of Theorem 2.153. Proposition 2.150 shows that there exist $q \in \mathbb{Z}$ and $r \in \{0,1,\ldots,N-1\}$ such that $n = qN + r$. Consider these q and r, and denote them by q_0 and r_0. Thus, $q_0 \in \mathbb{Z}$ and $r_0 \in \{0,1,\ldots,N-1\}$ and $n = q_0N + r_0$. From $q_0 \in \mathbb{Z}$ and $r_0 \in \{0,1,\ldots,N-1\}$, we obtain $(q_0,r_0) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$. Hence, there exists **at least one** pair $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$ (namely, $(q,r) = (q_0,r_0)$).

Now, let (q_1,r_1) and (q_2,r_2) be two pairs $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$. We shall prove that $(q_1,r_1) = (q_2,r_2)$.

We have assumed that (q_1,r_1) is a pair $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$. In other words, (q_1,r_1) is a pair in $\mathbb{Z} \times \{0,1,\ldots,N-1\}$ and satisfies $n = q_1N + r_1$. Similarly, (q_2,r_2) is a pair in $\mathbb{Z} \times \{0,1,\ldots,N-1\}$ and satisfies $n = q_2N + r_2$.

Hence, $q_1N + r_1 = n = q_2N + r_2$. Thus, Lemma 2.152 yields $(q_1,r_1) = (q_2,r_2)$.

Let us now forget that we fixed (q_1,r_1) and (q_2,r_2). We thus have shown that if (q_1,r_1) and (q_2,r_2) are two pairs $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$, then $(q_1,r_1) = (q_2,r_2)$. In other words, any two pairs $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$ must be equal. In other words, there exists **at most one** pair $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$. Since we also know that there exists **at least one** such pair, we can therefore conclude that there exists **exactly one** such pair. In other words, there is a unique pair $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$. This proves Theorem 2.153. $\square$

Definition 2.154. Let N be a positive integer. Let $n \in \mathbb{Z}$. Theorem 2.153 says that there is a unique pair $(q,r) \in \mathbb{Z} \times \{0,1,\ldots,N-1\}$ such that $n = qN + r$. Consider this pair (q,r). Then, q is called the *quotient of the division of n by N* (or the *quotient obtained when n is divided by N*), whereas r is called the *remainder of the division of n by N* (or the *remainder obtained when n is divided by N*).

For example, the quotient of the division of 7 by 3 is 2, whereas the remainder of the division of 7 by 3 is 1 (because $(2,1)$ is a pair in $\mathbb{Z} \times \{0,1,2\}$ such that $7 = 2 \cdot 3 + 1$).

We collect some basic properties of remainders:

Corollary 2.155. Let N be a positive integer. Let $n \in \mathbb{Z}$. Let $n\%N$ denote the remainder of the division of n by N.

(a) Then, $n\%N \in \{0,1,\ldots,N-1\}$ and $n\%N \equiv n \bmod N$.

(b) We have $N \mid n$ if and only if $n\%N = 0$.
(c) Let $c \in \{0, 1, \ldots, N-1\}$ be such that $c \equiv n \mod N$. Then, $c = n\%N$.

Proof of Corollary 2.155. Theorem 2.153 says that there is a unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$ such that $n = qN + r$. Consider this pair (q, r). Then, the remainder of the division of n by N is r (because this is how this remainder was defined). In other words, $n\%N$ is r (since $n\%N$ is the remainder of the division of n by N). Thus, $n\%N = r$. But $N \mid qN$ (since q is an integer), so that $qN \equiv 0 \mod N$. Hence, $\underbrace{qN}_{\equiv 0 \mod N} + r \equiv 0 + r = r \mod N$. Hence, $r \equiv qN + r = n \mod N$, so that $n\%N = r \equiv n \mod N$. Furthermore, $n\%N = r \in \{0, 1, \ldots, N-1\}$ (since $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$). This completes the proof of Corollary 2.155 **(a)**.

(b) We have the following implication:

$$(N \mid n) \Longrightarrow (n\%N = 0). \tag{211}$$

[*Proof of (211):* Assume that $N \mid n$. We must prove that $n\%N = 0$.

We have $N \mid n$. In other words, there exists some integer w such that $n = Nw$. Consider this w.

We have $N - 1 \in \mathbb{N}$ (since N is a positive integer), thus $0 \in \{0, 1, \ldots, N-1\}$. From $w \in \mathbb{Z}$ and $0 \in \{0, 1, \ldots, N-1\}$, we obtain $(w, 0) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$. Also, $wN + 0 = wN = Nw = n = qN + r$. Hence, Lemma 2.152 (applied to $(q_1, r_1) = (w, 0)$ and $(q_2, r_2) = (q, r)$) yields $(w, 0) = (q, r)$. In other words, $w = q$ and $0 = r$. Hence, $r = 0$, so that $n\%N = r = 0$. This proves the implication (211).]

Next, we have the following implication:

$$(n\%N = 0) \Longrightarrow (N \mid n). \tag{212}$$

[*Proof of (212):* Assume that $n\%N = 0$. We must prove that $N \mid n$.

We have $n = qN + \underbrace{r}_{= n\%N = 0} = qN$. Thus, $N \mid n$. This proves the implication (212).]

Combining the two implications (211) and (212), we obtain the logical equivalence $(N \mid n) \Longleftrightarrow (n\%N = 0)$. In other words, we have $N \mid n$ if and only if $n\%N = 0$. This proves Corollary 2.155 **(b)**.

(c) We have $c \equiv n \mod N$. In other words, $N \mid c - n$. In other words, there exists some integer w such that $c - n = Nw$. Consider this w.

From $-w \in \mathbb{Z}$ and $c \in \{0, 1, \ldots, N-1\}$, we obtain $(-w, c) \in \mathbb{Z} \times \{0, 1, \ldots, N-1\}$. Also, from $c - n = Nw$, we obtain $n = c - Nw = (-w) N + c$, so that $(-w) N + c = n = qN + r$. Hence, Lemma 2.152 (applied to $(q_1, r_1) = (-w, c)$ and $(q_2, r_2) = (q, r)$) yields $(-w, c) = (q, r)$. In other words, $-w = q$ and $c = r$. Hence, $c = r = n\%N$. This proves Corollary 2.155 **(c)**. □

Note that parts **(a)** and **(c)** of Corollary 2.155 (taken together) characterize the remainder $n\%N$ as the unique element of $\{0, 1, \ldots, N-1\}$ that is congruent to n modulo N. Corollary 2.155 **(b)** provides a simple algorithm to check whether a given integer n is divisible by a given positive integer N; namely, it suffices to compute the remainder $n\%N$ and check whether $n\%N = 0$.

Let us further illustrate the usefulness of Theorem 2.153 by proving a fundamental property of odd numbers. Recall the following standard definitions:

Definition 2.156. Let $n \in \mathbb{Z}$.

(a) We say that the integer n is *even* if and only if n is divisible by 2.

(b) We say that the integer n is *odd* if and only if n is not divisible by 2.

This definition shows that any integer n is either even or odd (but not both at the same time).

It is clear that an integer n is even if and only if it can be written in the form $n = 2m$ for some $m \in \mathbb{Z}$. Moreover, this m is unique (because $n = 2m$ implies $m = n/2$). Let us prove a similar property for odd numbers:

Proposition 2.157. Let $n \in \mathbb{Z}$.

(a) The integer n is odd if and only if n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$.

(b) This m is unique if it exists. (That is, any two integers $m \in \mathbb{Z}$ satisfying $n = 2m + 1$ must be equal.)

We shall use Theorem 2.153 several times in the below proof (far more than necessary), mostly to illustrate how it can be applied.

Proof of Proposition 2.157. **(a)** Let us first prove the logical implication

$$(n \text{ is odd}) \implies (\text{there exists an } m \in \mathbb{Z} \text{ such that } n = 2m + 1) . \qquad (213)$$

[*Proof of (213):* Assume that n is odd. We must prove that there exists an $m \in \mathbb{Z}$ such that $n = 2m + 1$.

Theorem 2.153 (applied to $N = 2$) yields that there is a unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2 - 1\}$ such that $n = q \cdot 2 + r$. Consider this (q, r). From $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2 - 1\}$, we obtain $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, 2 - 1\} = \{0, 1\}$.

We know that n is odd; in other words, n is not divisible by 2 (by the definition of "odd"). If we had $n = 2q$, then n would be divisible by 2, which would contradict the fact that n is not divisible by 2. Hence, we cannot have $n = 2q$. If we had $r = 0$, then we would have $n = \underbrace{q \cdot 2}_{=2q} + \underbrace{r}_{=0} = 2q$, which would contradict the fact that we cannot have $n = 2q$. Hence, we cannot have $r = 0$. Thus, $r \neq 0$.

Combining $r \in \{0, 1\}$ with $r \neq 0$, we obtain $r \in \{0, 1\} \setminus \{0\} = \{1\}$. Thus, $r = 1$. Hence, $n = \underbrace{q \cdot 2}_{=2q} + \underbrace{r}_{=1} = 2q + 1$. Thus, there exists an $m \in \mathbb{Z}$ such that $n = 2m + 1$ (namely, $m = q$). This proves the implication (213).]

Next, we shall prove the logical implication

$$(\text{there exists an } m \in \mathbb{Z} \text{ such that } n = 2m + 1) \implies (n \text{ is odd}) . \qquad (214)$$

[*Proof of (214):* Assume that there exists an $m \in \mathbb{Z}$ such that $n = 2m + 1$. We must prove that n is odd.

We have assumed that there exists an $m \in \mathbb{Z}$ such that $n = 2m + 1$. Consider this m. Thus, the pair $(m, 1)$ belongs to $\mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ (since $m \in \mathbb{Z}$ and $1 \in \{0, 1, \ldots, 2-1\}$) and satisfies $n = m \cdot 2 + 1$ (since $n = \underbrace{2m}_{=m \cdot 2} + 1 = m \cdot 2 + 1$). In other words, the pair $(m, 1)$ is a pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$.

Now, assume (for the sake of contradiction) that n is divisible by 2. Thus, there exists some integer w such that $n = 2w$. Consider this w. Thus, the pair $(w, 0)$ belongs to $\mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ (since $w \in \mathbb{Z}$ and $0 \in \{0, 1, \ldots, 2-1\}$) and satisfies $n = w \cdot 2 + 0$ (since $n = 2w = w \cdot 2 = w \cdot 2 + 0$). In other words, the pair $(w, 0)$ is a pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$.

Theorem 2.153 (applied to $N = 2$) yields that there is a unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$. Thus, there exists **at most** one such pair. In other words, any two such pairs must be equal. Hence, the two pairs $(m, 1)$ and $(w, 0)$ must be equal (since $(m, 1)$ and $(w, 0)$ are two pairs $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$). In other words, $(m, 1) = (w, 0)$. In other words, $m = w$ and $1 = 0$. But $1 = 0$ is clearly absurd. Thus, we have obtained a contradiction. This shows that our assumption (that n is divisible by 2) was wrong. Hence, n is not divisible by 2. In other words, n is odd (by the definition of "odd"). This proves the implication (214).]

Combining the two implications (213) and (214), we obtain the logical equivalence

$$\begin{aligned}(n \text{ is odd}) &\iff (\text{there exists an } m \in \mathbb{Z} \text{ such that } n = 2m + 1) \\ &\iff (n \text{ can be written in the form } n = 2m + 1 \text{ for some } m \in \mathbb{Z}).\end{aligned}$$

In other words, the integer n is odd if and only if n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$. This proves Proposition 2.157 **(a)**.

(b) This is easy to prove in any way, but let us prove this using Theorem 2.153 just in order to illustrate the use of the latter theorem.

We must prove that any two integers $m \in \mathbb{Z}$ satisfying $n = 2m + 1$ must be equal.

Let m_1 and m_2 be two integers $m \in \mathbb{Z}$ satisfying $n = 2m + 1$. We shall show that $m_1 = m_2$.

We know that m_1 is an integer $m \in \mathbb{Z}$ satisfying $n = 2m + 1$. In other words, m_1 is an integer in $\mathbb{Z}$ and satisfies $n = 2m_1 + 1$. Thus, the pair $(m_1, 1)$ belongs to $\mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ (since $m_1 \in \mathbb{Z}$ and $1 \in \{0, 1, \ldots, 2-1\}$) and satisfies $n = m_1 \cdot 2 + 1$ (since $n = \underbrace{2m_1}_{=m_1 \cdot 2} + 1 = m_1 \cdot 2 + 1$). In other words, the pair $(m_1, 1)$ is a pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$. The same argument (applied to m_2 instead of m_1) shows that $(m_2, 1)$ is a pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$.

Theorem 2.153 (applied to $N = 2$) yields that there is a unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$. Thus, there exists **at most** one such pair. In other words, any two such pairs must be equal. Hence, the two pairs $(m_1, 1)$ and $(m_2, 1)$ must be equal (since $(m_1, 1)$ and $(m_2, 1)$ are two pairs $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, 2-1\}$ such that $n = q \cdot 2 + r$). In other words, $(m_1, 1) = (m_2, 1)$. In other words, $m_1 = m_2$ and $1 = 1$. Hence, we have shown that $m_1 = m_2$.

Now, forget that we fixed m_1 and m_2. We thus have proven that if m_1 and m_2 are two integers $m \in \mathbb{Z}$ satisfying $n = 2m + 1$, then $m_1 = m_2$. In other words, any two integers $m \in \mathbb{Z}$ satisfying $n = 2m + 1$ must be equal. In other words, the m in Proposition 2.157 **(a)** is unique. This proves Proposition 2.157 **(b)**. □

We can use this to obtain the following fundamental fact:

Corollary 2.158. Let $n \in \mathbb{Z}$.

(a) If n is even, then $(-1)^n = 1$.

(b) If n is odd, then $(-1)^n = -1$.

Proof of Corollary 2.158. **(a)** Assume that n is even. In other words, n is divisible by 2 (by the definition of "even"). In other words, $2 \mid n$. In other words, there exists an integer w such that $n = 2w$. Consider this w. From $n = 2w$, we obtain $(-1)^n = (-1)^{2w} = \left(\underbrace{(-1)^2}_{=1} \right)^w = 1^w = 1$. This proves Corollary 2.158 **(a)**.

(b) Assume that n is odd. Proposition 2.157 **(a)** shows that the integer n is odd if and only if n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$. Hence, n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$ (since the integer n is odd). Consider this m. From $n = 2m + 1$, we obtain

$$(-1)^n = (-1)^{2m+1} = (-1)^{2m} (-1) = - \underbrace{(-1)^{2m}}_{=\left((-1)^2\right)^m} = - \left(\underbrace{(-1)^2}_{=1} \right)^m = - \underbrace{1^m}_{=1} = -1.$$

This proves Corollary 2.158 **(b)**. □

Let us state two more fundamental facts, which are proven in Exercise 2.4:

Proposition 2.159. Let u and v be two integers. Then, we have the following chain of logical equivalences:

$$\begin{aligned} (u \equiv v \bmod 2) &\iff (u \text{ and } v \text{ are either both even or both odd}) \\ &\iff \left((-1)^u = (-1)^v\right). \end{aligned}$$

Proposition 2.160. Let $n \in \mathbb{Z}$.

(a) The integer n is even if and only if $n \equiv 0 \bmod 2$.

(b) The integer n is odd if and only if $n \equiv 1 \bmod 2$.

Exercise 2.4. Prove Proposition 2.159 and Proposition 2.160.

Exercise 2.5. Let N be a positive integer. Let $p \in \mathbb{Z}$ and $h \in \mathbb{Z}$. Prove that there exists a **unique** element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$.

Exercise 2.6. Let $k \in \mathbb{N}$, $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$.
(a) Prove that $a - b \mid a^k - b^k$.
(b) Assume that k is odd. Prove that $a + b \mid a^k + b^k$.

Exercise 2.7. Fix an **odd** positive integer r. Consider the sequence $(b_0, b_1, b_2, \ldots)$ defined in Proposition 2.66.

Prove that $b_n + 1 \mid b_{n-1}(b_{n+2} + 1)$ for each positive integer n. (Note that the statement "$b_n + 1 \mid b_{n-1}(b_{n+2} + 1)$" makes sense, since Proposition 2.66 **(a)** yields that all three numbers b_n, b_{n-1}, b_{n+2} belong to $\mathbb{N}$.)

2.15.3. Backwards induction principles

When we use Theorem 2.149 to prove a statement, we can regard the proof of Assumption 2 as a (regular) induction step ("forwards induction step"), and regard the proof of Assumption 3 as a sort of "backwards induction step". There are also "backwards induction principles" which include a "backwards induction step" but no "forwards induction step". Here are two such principles:

Theorem 2.161. Let $g \in \mathbb{Z}$. Let $\mathbb{Z}_{\leq g}$ be the set $\{g, g-1, g-2, \ldots\}$ (that is, the set of all integers that are $\leq g$). For each $n \in \mathbb{Z}_{\leq g}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: The statement $\mathcal{A}(g)$ holds.

Assumption 2: If $m \in \mathbb{Z}_{\leq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m-1)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\leq g}$.

Theorem 2.162. Let $g \in \mathbb{Z}$ and $h \in \mathbb{Z}$. For each $n \in \{g, g+1, \ldots, h\}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: If $g \leq h$, then the statement $\mathcal{A}(h)$ holds.

Assumption 2: If $m \in \{g+1, g+2, \ldots, h\}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m-1)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$.

Theorem 2.161 is an analogue of Theorem 2.53, while Theorem 2.162 is an analogue of Theorem 2.74. However, it is not hard to derive these theorems from the induction principles we already know:

Exercise 2.8. Prove Theorem 2.161 and Theorem 2.162.

It is also easy to state and prove a "backwards" analogue of Theorem 2.60. (We leave this to the reader.)

A proof using Theorem 2.161 or using Theorem 2.162 is usually called a *proof by descending induction* or a *proof by backwards induction*.

2.16. Induction from $k-1$ to k

2.16.1. The principle

Let us next show yet another "alternative induction principle", which differs from Theorem 2.53 in a mere notational detail:

Theorem 2.163. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{A}(n)$ be a logical statement.

Assume the following:

Assumption 1: The statement $\mathcal{A}(g)$ holds.

Assumption 2: If $k \in \mathbb{Z}_{\geq g+1}$ is such that $\mathcal{A}(k-1)$ holds, then $\mathcal{A}(k)$ also holds.

Then, $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Roughly speaking, this Theorem 2.163 is just Theorem 2.53, except that the variable m in Assumption 2 has been renamed as $k-1$. Consequently, it stands to reason that Theorem 2.163 can easily be derived from Theorem 2.53. Here is the derivation in full detail:

Proof of Theorem 2.163. For each $n \in \mathbb{Z}_{\geq g}$, we define the logical statement $\mathcal{B}(n)$ to be the statement $\mathcal{A}(n)$. Thus, $\mathcal{B}(n) = \mathcal{A}(n)$ for each $n \in \mathbb{Z}_{\geq g}$. Applying this to $n = g$, we obtain $\mathcal{B}(g) = \mathcal{A}(g)$ (since $g \in \mathbb{Z}_{\geq g}$).

We shall now show that the two Assumptions A and B of Corollary 2.61 are satisfied.

Indeed, recall that Assumption 1 is satisfied. In other words, the statement $\mathcal{A}(g)$ holds. In other words, the statement $\mathcal{B}(g)$ holds (since $\mathcal{B}(g) = \mathcal{A}(g)$). In other words, Assumption A is satisfied.

We shall next show that Assumption B is satisfied. Indeed, let $p \in \mathbb{Z}_{\geq g}$ be such that $\mathcal{B}(p)$ holds. Recall that the statement $\mathcal{B}(p)$ was defined to be the statement $\mathcal{A}(p)$. Thus, $\mathcal{B}(p) = \mathcal{A}(p)$. Hence, $\mathcal{A}(p)$ holds (since $\mathcal{B}(p)$ holds). Now, let $k = p+1$. We know that $p \in \mathbb{Z}_{\geq g}$; in other words, p is an integer and satisfies

$p \geq g$. Hence, $k = p + 1$ is an integer as well and satisfies $k = \underbrace{p}_{\geq g} + 1 \geq g + 1$. In other words, $k \in \mathbb{Z}_{\geq g+1}$. Moreover, from $k = p + 1$, we obtain $k - 1 = p$. Hence, $\mathcal{A}(k-1) = \mathcal{A}(p)$. Thus, $\mathcal{A}(k-1)$ holds (since $\mathcal{A}(p)$ holds). Thus, Assumption 2 shows that $\mathcal{A}(k)$ also holds. But the statement $\mathcal{B}(k)$ was defined to be the statement $\mathcal{A}(k)$. Hence, $\mathcal{B}(k) = \mathcal{A}(k)$, so that $\mathcal{A}(k) = \mathcal{B}(k) = \mathcal{B}(p+1)$ (since $k = p + 1$). Thus, the statement $\mathcal{B}(p+1)$ holds (since $\mathcal{A}(k)$ holds). Now, forget that we fixed p. We thus have shown that if $p \in \mathbb{Z}_{\geq g}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds. In other words, Assumption B is satisfied.

We have now proven that both Assumptions A and B of Corollary 2.61 are satisfied. Hence, Corollary 2.61 shows that $\mathcal{B}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$. In other words, $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$ (because each $n \in \mathbb{Z}_{\geq g}$ satisfies $\mathcal{B}(n) = \mathcal{A}(n)$ (by the definition of $\mathcal{B}(n)$)). This proves Theorem 2.163. □

Proofs that use Theorem 2.163 are usually called *proofs by induction* or *induction proofs*. As an example of such a proof, let us show the following identity:

Proposition 2.164. For every $n \in \mathbb{N}$, we have

$$\sum_{i=1}^{2n} \frac{(-1)^{i-1}}{i} = \sum_{i=n+1}^{2n} \frac{1}{i}. \tag{215}$$

The equality (215) can be rewritten as

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \cdots + \frac{1}{2n-1} - \frac{1}{2n} = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$$

(where all the signs on the right hand side are $+$ signs, whereas the signs on the left hand side alternate between $+$ signs and $-$ signs).

Proof of Proposition 2.164. For each $n \in \mathbb{Z}_{\geq 0}$, we let $\mathcal{A}(n)$ be the statement

$$\left(\sum_{i=1}^{2n} \frac{(-1)^{i-1}}{i} = \sum_{i=n+1}^{2n} \frac{1}{i} \right).$$

Our next goal is to prove the statement $\mathcal{A}(n)$ for each $n \in \mathbb{Z}_{\geq 0}$.

We first notice that the statement $\mathcal{A}(0)$ holds[96].

[96] *Proof.* We have $\sum_{i=1}^{2\cdot 0} \frac{(-1)^{i-1}}{i} = (\text{empty sum}) = 0$. Comparing this with $\sum_{i=0+1}^{2\cdot 0} \frac{1}{i} = (\text{empty sum}) = 0$, we obtain $\sum_{i=1}^{2\cdot 0} \frac{(-1)^{i-1}}{i} = \sum_{i=0+1}^{2\cdot 0} \frac{1}{i}$. But this is precisely the statement $\mathcal{A}(0)$ (since $\mathcal{A}(0)$ is defined to be the statement $\left(\sum_{i=1}^{2\cdot 0} \frac{(-1)^{i-1}}{i} = \sum_{i=0+1}^{2\cdot 0} \frac{1}{i} \right)$). Hence, the statement $\mathcal{A}(0)$ holds.

Now, we claim that

$$\text{if } k \in \mathbb{Z}_{\geq 0+1} \text{ is such that } \mathcal{A}(k-1) \text{ holds, then } \mathcal{A}(k) \text{ also holds.} \qquad (216)$$

[*Proof of (216):* Let $k \in \mathbb{Z}_{\geq 0+1}$ be such that $\mathcal{A}(k-1)$ holds. We must show that $\mathcal{A}(k)$ also holds.

We have $k \in \mathbb{Z}_{\geq 0+1}$. Thus, k is an integer and satisfies $k \geq 0+1 = 1$.

We have assumed that $\mathcal{A}(k-1)$ holds. In other words,

$$\sum_{i=1}^{2(k-1)} \frac{(-1)^{i-1}}{i} = \sum_{i=(k-1)+1}^{2(k-1)} \frac{1}{i} \qquad (217)$$

holds[97].

We have $(-1)^{2(k-1)} = \left(\underbrace{(-1)^2}_{=1}\right)^{k-1} = 1^{k-1} = 1$. But $2k-1 = 2(k-1)+1$. Thus, $(-1)^{2k-1} = (-1)^{2(k-1)+1} = \underbrace{(-1)^{2(k-1)}}_{=1} \underbrace{(-1)^1}_{=-1} = -1$.

Now, $k \geq 1$, so that $2k \geq 2$ and therefore $2k-1 \geq 1$. Hence, we can split off the addend for $i = 2k-1$ from the sum $\sum_{i=1}^{2k-1} \frac{(-1)^{i-1}}{i}$. We thus obtain

$$\begin{aligned}
\sum_{i=1}^{2k-1} \frac{(-1)^{i-1}}{i} &= \sum_{i=1}^{(2k-1)-1} \frac{(-1)^{i-1}}{i} + \frac{(-1)^{(2k-1)-1}}{2k-1} \\
&= \underbrace{\sum_{i=1}^{2(k-1)} \frac{(-1)^{i-1}}{i}}_{\substack{= \sum_{i=(k-1)+1}^{2(k-1)} \frac{1}{i} \\ \text{(by (217))}}} + \underbrace{\frac{(-1)^{2(k-1)}}{2k-1}}_{\substack{= \frac{1}{2k-1} \\ \text{(since } (-1)^{2(k-1)} = 1)}} \\
&\qquad (\text{since } (2k-1)-1 = 2(k-1)) \\
&= \sum_{i=(k-1)+1}^{2(k-1)} \frac{1}{i} + \frac{1}{2k-1} = \sum_{i=k}^{2k-2} \frac{1}{i} + \frac{1}{2k-1}
\end{aligned} \qquad (218)$$

(since $(k-1)+1 = k$ and $2(k-1) = 2k-2$).

On the other hand, $2k \geq 2 \geq 1$. Hence, we can split off the addend for $i = 2k$

[97] because $\mathcal{A}(k-1)$ is defined to be the statement $\left(\sum_{i=1}^{2(k-1)} \frac{(-1)^{i-1}}{i} = \sum_{i=(k-1)+1}^{2(k-1)} \frac{1}{i}\right)$

from the sum $\sum\limits_{i=1}^{2k} \dfrac{(-1)^{i-1}}{i}$. We thus obtain

$$\sum_{i=1}^{2k} \frac{(-1)^{i-1}}{i} = \underbrace{\sum_{i=1}^{2k-1} \frac{(-1)^{i-1}}{i}}_{\substack{= \sum\limits_{i=k}^{2k-2} \frac{1}{i} + \frac{1}{2k-1} \\ \text{(by (218))}}} + \underbrace{\frac{(-1)^{2k-1}}{2k}}_{\substack{= \frac{-1}{2k} \\ \text{(since } (-1)^{2k-1} = -1)}}$$

$$= \sum_{i=k}^{2k-2} \frac{1}{i} + \frac{1}{2k-1} + \frac{-1}{2k}. \tag{219}$$

But we have $(2k-1) - k = k - 1 \geq 0$ (since $k \geq 1$). Thus, $2k - 1 \geq k$. Thus, we can split off the addend for $i = 2k-1$ from the sum $\sum\limits_{i=k}^{2k-1} \dfrac{1}{i}$. We thus obtain

$$\sum_{i=k}^{2k-1} \frac{1}{i} = \sum_{i=k}^{(2k-1)-1} \frac{1}{i} + \frac{1}{2k-1} = \sum_{i=k}^{2k-2} \frac{1}{i} + \frac{1}{2k-1} \tag{220}$$

(since $(2k-1) - 1 = 2k-2$). Hence, (219) becomes

$$\sum_{i=1}^{2k} \frac{(-1)^{i-1}}{i} = \underbrace{\sum_{i=k}^{2k-2} \frac{1}{i} + \frac{1}{2k-1}}_{\substack{= \sum\limits_{i=k}^{2k-1} \frac{1}{i} \\ \text{(by (220))}}} + \frac{-1}{2k} = \sum_{i=k}^{2k-1} \frac{1}{i} + \frac{-1}{2k}. \tag{221}$$

But we have $k + 1 \leq 2k$ (since $2k - (k+1) = k - 1 \geq 0$). Thus, we can split off the addend for $i = 2k$ from the sum $\sum\limits_{i=k+1}^{2k} \dfrac{1}{i}$. We thus obtain

$$\sum_{i=k+1}^{2k} \frac{1}{i} = \sum_{i=k+1}^{2k-1} \frac{1}{i} + \frac{1}{2k}.$$

Hence,

$$\sum_{i=k+1}^{2k-1} \frac{1}{i} = \sum_{i=k+1}^{2k} \frac{1}{i} - \frac{1}{2k}. \tag{222}$$

Also, $k \leq 2k - 1$ (since $(2k-1) - k = k - 1 \geq 0$). Thus, we can split off the addend for $i = k$ from the sum $\sum\limits_{i=k}^{2k-1} \dfrac{1}{i}$. We thus obtain

$$\sum_{i=k}^{2k-1} \frac{1}{i} = \frac{1}{k} + \underbrace{\sum_{i=k+1}^{2k-1} \frac{1}{i}}_{\substack{= \sum\limits_{i=k+1}^{2k} \frac{1}{i} - \frac{1}{2k} \\ \text{(by (222))}}} = \frac{1}{k} + \sum_{i=k+1}^{2k} \frac{1}{i} - \frac{1}{2k} = \sum_{i=k+1}^{2k} \frac{1}{i} + \underbrace{\frac{1}{k} - \frac{1}{2k}}_{= \frac{1}{2k}} = \sum_{i=k+1}^{2k} \frac{1}{i} + \frac{1}{2k}.$$

Subtracting $\frac{1}{2k}$ from this equality, we obtain

$$\sum_{i=k}^{2k-1} \frac{1}{i} - \frac{1}{2k} = \sum_{i=k+1}^{2k} \frac{1}{i}.$$

Hence,

$$\sum_{i=k+1}^{2k} \frac{1}{i} = \sum_{i=k}^{2k-1} \frac{1}{i} - \frac{1}{2k} = \sum_{i=k}^{2k-1} \frac{1}{i} + \frac{-1}{2k}.$$

Comparing this with (221), we obtain

$$\sum_{i=1}^{2k} \frac{(-1)^{i-1}}{i} = \sum_{i=k+1}^{2k} \frac{1}{i}. \tag{223}$$

But this is precisely the statement $\mathcal{A}(k)$ [98]. Thus, the statement $\mathcal{A}(k)$ holds.

Now, forget that we fixed k. We thus have shown that if $k \in \mathbb{Z}_{\geq 0+1}$ is such that $\mathcal{A}(k-1)$ holds, then $\mathcal{A}(k)$ also holds. This proves (216).]

Now, both assumptions of Theorem 2.163 (applied to $g = 0$) are satisfied (indeed, Assumption 1 holds because the statement $\mathcal{A}(0)$ holds, whereas Assumption 2 holds because of (216)). Thus, Theorem 2.163 (applied to $g = 0$) shows that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq 0}$. In other words, $\sum_{i=1}^{2n} \frac{(-1)^{i-1}}{i} = \sum_{i=n+1}^{2n} \frac{1}{i}$ holds for each $n \in \mathbb{Z}_{\geq 0}$ (since $\mathcal{A}(n)$ is the statement $\left(\sum_{i=1}^{2n} \frac{(-1)^{i-1}}{i} = \sum_{i=n+1}^{2n} \frac{1}{i}\right)$). In other words, $\sum_{i=1}^{2n} \frac{(-1)^{i-1}}{i} = \sum_{i=n+1}^{2n} \frac{1}{i}$ holds for each $n \in \mathbb{N}$ (because $\mathbb{Z}_{\geq 0} = \mathbb{N}$). This proves Proposition 2.164. □

2.16.2. Conventions for writing proofs using "$k-1$ to k" induction

Just like most of the other induction principles that we have so far introduced, Theorem 2.163 is not usually invoked explicitly when it is used; instead, its use is signalled by certain words:

Convention 2.165. Let $g \in \mathbb{Z}$. For each $n \in \mathbb{Z}_{\geq g}$, let $\mathcal{A}(n)$ be a logical statement. Assume that you want to prove that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\geq g}$.

Theorem 2.163 offers the following strategy for proving this: First show that Assumption 1 of Theorem 2.163 is satisfied; then, show that Assumption 2 of Theorem 2.163 is satisfied; then, Theorem 2.163 automatically completes your proof.

[98] because $\mathcal{A}(k)$ is defined to be the statement $\left(\sum_{i=1}^{2k} \frac{(-1)^{i-1}}{i} = \sum_{i=k+1}^{2k} \frac{1}{i}\right)$

A proof that follows this strategy is called a *proof by induction on n* (or *proof by induction over n*) *starting at g* or (less precisely) an *inductive proof*. Most of the time, the words "starting at g" are omitted, since the value of g is usually clear from the statement that is being proven. Usually, the statements $\mathcal{A}(n)$ are not explicitly stated in the proof either, since they can also be inferred from the context.

The proof that Assumption 1 is satisfied is called the *induction base* (or *base case*) of the proof. The proof that Assumption 2 is satisfied is called the *induction step* of the proof.

In order to prove that Assumption 2 is satisfied, you will usually want to fix a $k \in \mathbb{Z}_{\geq g+1}$ such that $\mathcal{A}(k-1)$ holds, and then prove that $\mathcal{A}(k)$ holds. In other words, you will usually want to fix $k \in \mathbb{Z}_{\geq g+1}$, assume that $\mathcal{A}(k-1)$ holds, and then prove that $\mathcal{A}(k)$ holds. When doing so, it is common to refer to the assumption that $\mathcal{A}(k-1)$ holds as the *induction hypothesis* (or *induction assumption*).

This language is exactly the same that was introduced in Convention 2.56 for proofs by "standard" induction starting at g. The only difference between proofs that use Theorem 2.53 and proofs that use Theorem 2.163 is that the induction step in the former proofs assumes $\mathcal{A}(m)$ and proves $\mathcal{A}(m+1)$, whereas the induction step in the latter proofs assumes $\mathcal{A}(k-1)$ and proves $\mathcal{A}(k)$. (Of course, the letters "m" and "k" are not set in stone; any otherwise unused letters can be used in their stead. Thus, what distinguishes proofs that use Theorem 2.53 from proofs that use Theorem 2.163 is not the letter they use, but the "$+1$" versus the "-1".)

Let us repeat the above proof of Proposition 2.164 (or, more precisely, its non-computational part) using this language:

Proof of Proposition 2.164 (second version). We must prove (215) for every $n \in \mathbb{N}$. In other words, we must prove (215) for every $n \in \mathbb{Z}_{\geq 0}$ (since $\mathbb{N} = \mathbb{Z}_{\geq 0}$). We shall prove this by induction on n starting at 0:

Induction base: We have $\sum\limits_{i=1}^{2\cdot 0} \dfrac{(-1)^{i-1}}{i} = (\text{empty sum}) = 0$. Comparing this with $\sum\limits_{i=0+1}^{2\cdot 0} \dfrac{1}{i} = (\text{empty sum}) = 0$, we obtain $\sum\limits_{i=1}^{2\cdot 0} \dfrac{(-1)^{i-1}}{i} = \sum\limits_{i=0+1}^{2\cdot 0} \dfrac{1}{i}$. In other words, (215) holds for $n = 0$. This completes the induction base.

Induction step: Let $k \in \mathbb{Z}_{\geq 1}$. Assume that (215) holds for $n = k-1$. We must show that (215) holds for $n = k$.

We have $k \in \mathbb{Z}_{\geq 1}$. In other words, k is an integer and satisfies $k \geq 1$.

We have assumed that (215) holds for $n = k-1$. In other words,

$$\sum_{i=1}^{2(k-1)} \frac{(-1)^{i-1}}{i} = \sum_{i=(k-1)+1}^{2(k-1)} \frac{1}{i}. \tag{224}$$

From here, we can obtain

$$\sum_{i=1}^{2k} \frac{(-1)^{i-1}}{i} = \sum_{i=k+1}^{2k} \frac{1}{i}. \tag{225}$$

(Indeed, we can derive (225) from (224) in exactly the same way as we derived (223) from (217) in the above first version of the proof of Proposition 2.164; nothing about this argument needs to be changed, so we have no reason to repeat it.)

But the equality (225) shows that (215) holds for $n = k$. This completes the induction step. Hence, (215) is proven by induction. This proves Proposition 2.164. □

Exercise 2.9. Let $n \in \mathbb{N}$. Prove that

$$\sum_{k=0}^{n} (-1)^k (k+1) = \begin{cases} n/2+1, & \text{if } n \text{ is even;} \\ -(n+1)/2, & \text{if } n \text{ is odd} \end{cases}.$$

The claim of Exercise 2.9 can be rewritten as

$$1 - 2 + 3 - 4 \pm \cdots + (-1)^n (n+1) = \begin{cases} n/2+1, & \text{if } n \text{ is even;} \\ -(n+1)/2, & \text{if } n \text{ is odd} \end{cases}.$$

3. On binomial coefficients

The present chapter is about *binomial coefficients*. They are used in almost every part of mathematics, and studying them provides good opportunities to practice the arts of mathematical induction and of finding combinatorial bijections.

Identities involving binomial coefficients are legion, and books have been written about them (let me mention [GrKnPa94, Chapter 5] as a highly readable introduction; but, e.g., Henry W. Gould's website goes far deeper down the rabbit hole). We shall only study a few of these identities.

3.1. Definitions and basic properties

3.1.1. The definition

Let us first define binomial coefficients:

Definition 3.1. Let $n \in \mathbb{N}$ and $m \in \mathbb{Q}$. Recall that $n!$ is a positive integer; thus, $n! \neq 0$. (Keep in mind that $0! = 1$.)

We define a rational number $\binom{m}{n}$ by

$$\binom{m}{n} = \frac{m(m-1)\cdots(m-n+1)}{n!}. \tag{226}$$

(This fraction is well-defined, since $n! \neq 0$. When $n = 0$, the numerator of this fraction (i.e., the product $m(m-1)\cdots(m-n+1)$) is an empty product. Recall that, by convention, an empty product is always defined to be 1.)

This number $\binom{m}{n}$ is called a *binomial coefficient*, and is often pronounced "m choose n".

We can extend this definition to the case when $m \in \mathbb{R}$ or $m \in \mathbb{C}$ (rather than $m \in \mathbb{Q}$) by using the same equality (226). Of course, in that case, $\binom{m}{n}$ will not be a rational number anymore.

Example 3.2. The formula (226) yields

$$\binom{4}{2} = \frac{4(4-1)}{2!} = \frac{4(4-1)}{2} = 6;$$
$$\binom{5}{1} = \frac{5}{1!} = \frac{5}{1} = 5;$$
$$\binom{8}{3} = \frac{8(8-1)(8-2)}{3!} = \frac{8(8-1)(8-2)}{6} = 56.$$

Here is a table of the binomial coefficients $\binom{n}{k}$ for all values $n \in \{-3, -2, -1, \ldots, 6\}$ and some of the values $k \in \{0, 1, 2, 3, 4, 5\}$. In the following table, each row corresponds to a value of n, while each southwest-northeast diagonal corresponds to a value of k:

											$k=0$ ↙		$k=1$ ↙		$k=2$ ↙		$k=3$ ↙
$n=-3$ →										1		-3		6		-10	
$n=-2$ →									1		-2		3		-4		
$n=-1$ →								1		-1		1		-1		1	
$n=0$ →							1		0		0		0		0		
$n=1$ →						1		1		0		0		0		0	
$n=2$ →					1		2		1		0		0		0		
$n=3$ →				1		3		3		1		0		0		0	
$n=4$ →			1		4		6		4		1		0		0		
$n=5$ →		1		5		10		10		5		1		0		0	
$n=6$ →	1		6		15		20		15		6		1		0		

The binomial coefficients $\binom{m}{n}$ form the so-called *Pascal's triangle*[99]. Let us state

[99]More precisely, the numbers $\binom{m}{n}$ for $m \in \mathbb{N}$ and $n \in \{0, 1, \ldots, m\}$ form Pascal's triangle. Nev-

a few basic properties of these numbers:

3.1.2. Simple formulas

Proposition 3.3. Let $m \in \mathbb{Q}$.

(a) We have

$$\binom{m}{0} = 1. \tag{227}$$

(b) We have

$$\binom{m}{1} = m. \tag{228}$$

Proof of Proposition 3.3. **(a)** The definition of $\binom{m}{0}$ yields

$$\binom{m}{0} = \frac{m(m-1)\cdots(m-0+1)}{0!}.$$

Since $m(m-1)\cdots(m-0+1) = (\text{a product of 0 integers}) = 1$, this rewrites as $\binom{m}{0} = \frac{1}{0!} = 1$ (since $0! = 1$). This proves Proposition 3.3 **(a)**.

(b) The definition of $\binom{m}{1}$ yields

$$\binom{m}{1} = \frac{m(m-1)\cdots(m-1+1)}{1!}.$$

Since $m(m-1)\cdots(m-1+1) = m$ (because the product $m(m-1)\cdots(m-1+1)$ consists of 1 factor only, and this factor is m), this rewrites as $\binom{m}{1} = \frac{m}{1!} = m$ (since $1! = 1$). This proves Proposition 3.3 **(b)**. □

Proposition 3.4. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$ be such that $m \geq n$. Then,

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}. \tag{229}$$

Remark 3.5. Caution: The formula (229) holds only for $m \in \mathbb{N}$ and $n \in \mathbb{N}$ satisfying $m \geq n$. Thus, neither $\binom{-3}{2}$ nor $\binom{1/3}{3}$ nor $\binom{2}{5}$ can be computed using this formula! Definition 3.1 thus can be used to compute $\binom{m}{n}$ in many more cases than (229) does.

ertheless, the "other" binomial coefficients (particularly the ones where m is a negative integer) are highly useful, too.

Proof of Proposition 3.4. Multiplying both sides of the equality (226) with $n!$, we obtain

$$n! \cdot \binom{m}{n} = m(m-1)\cdots(m-n+1). \tag{230}$$

But

$$\begin{aligned} m! &= m(m-1)\cdots 1 = (m(m-1)\cdots(m-n+1)) \cdot \underbrace{((m-n)(m-n-1)\cdots 1)}_{=(m-n)!} \\ &= (m(m-1)\cdots(m-n+1)) \cdot (m-n)!, \end{aligned}$$

so that $\dfrac{m!}{(m-n)!} = m(m-1)\cdots(m-n+1)$. Comparing this with (230), we obtain $n! \cdot \binom{m}{n} = \dfrac{m!}{(m-n)!}$. Dividing this equality by $n!$, we obtain $\binom{m}{n} = \dfrac{m!}{n!(m-n)!}$. Thus, Proposition 3.4 is proven. □

Proposition 3.6. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$ be such that $m < n$. Then,

$$\binom{m}{n} = 0. \tag{231}$$

Remark 3.7. Caution: The formula (231) is not true if we drop the condition $m \in \mathbb{N}$. For example, $\binom{-3}{2} = 6 \neq 0$ despite $-3 < 2$.

Proof of Proposition 3.6. We have $m \geq 0$ (since $m \in \mathbb{N}$). Also, $m < n$, so that $m \leq n-1$ (since m and n are integers). Thus, $m \in \{0, 1, \ldots, n-1\}$. Hence, $m - m$ is one of the n integers $m, m-1, \ldots, m-n+1$. Thus, one of the n factors of the product $m(m-1)\cdots(m-n+1)$ is $m - m = 0$. Therefore, the whole product $m(m-1)\cdots(m-n+1)$ is 0 (because if one of the factors of a product is 0, then the whole product must be 0). Thus, $m(m-1)\cdots(m-n+1) = 0$. Hence, (226) becomes

$$\begin{aligned} \binom{m}{n} &= \frac{m(m-1)\cdots(m-n+1)}{n!} = \frac{0}{n!} \qquad (\text{since } m(m-1)\cdots(m-n+1) = 0) \\ &= 0. \end{aligned}$$

This proves Proposition 3.6. □

Proposition 3.8. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$ be such that $m \geq n$. Then,

$$\binom{m}{n} = \binom{m}{m-n}. \tag{232}$$

Proposition 3.8 is commonly known as the *symmetry identity for the binomial coefficients*. Notice that Proposition 3.8 becomes false (and, with our definitions, actually meaningless) if the requirement that $m \in \mathbb{N}$ is dropped.

Proof of Proposition 3.8. We have $m - n \in \mathbb{N}$ (since $m \geq n$) and $m \geq m - n$ (since $n \geq 0$ (since $n \in \mathbb{N}$)). Hence, (229) (applied to $m - n$ instead of n) yields

$$\binom{m}{m-n} = \frac{m!}{(m-n)!\,(m-(m-n))!} = \frac{m!}{(m-(m-n))!\,(m-n)!} = \frac{m!}{n!\,(m-n)!}$$

(since $m - (m-n) = n$). Compared with (229), this yields $\binom{m}{n} = \binom{m}{m-n}$. Proposition 3.8 is thus proven. □

Proposition 3.9. Let $m \in \mathbb{N}$. Then,

$$\binom{m}{m} = 1. \tag{233}$$

Proof of Proposition 3.9. The equality (232) (applied to $n = m$) yields $\binom{m}{m} = \binom{m}{m-m} = \binom{m}{0} = 1$ (according to (227)). This proves Proposition 3.9. □

Exercise 3.1. Let $m \in \mathbb{N}$ and $(k_1, k_2, \ldots, k_m) \in \mathbb{N}^m$. Prove that $\frac{(k_1 + k_2 + \cdots + k_m)!}{k_1! k_2! \cdots k_m!}$ is a positive integer.

Remark 3.10. Let $m \in \mathbb{N}$ and $(k_1, k_2, \ldots, k_m) \in \mathbb{N}^m$. Exercise 3.1 shows that $\frac{(k_1 + k_2 + \cdots + k_m)!}{k_1! k_2! \cdots k_m!}$ is a positive integer. This positive integer is called a *multinomial coefficient*, and is often denoted by $\binom{n}{k_1, k_2, \ldots, k_m}$, where $n = k_1 + k_2 + \cdots + k_m$. (We shall avoid this particular notation, since it makes the meaning of $\binom{n}{k}$ slightly ambiguous: It could mean both the binomial coefficient $\binom{n}{k}$ and the multinomial coefficient $\binom{n}{k_1, k_2, \ldots, k_m}$ for $(k_1, k_2, \ldots, k_m) = (k)$. Fortunately, the ambiguity is not really an issue, because the only situation in which both meanings make sense is when $k = n \in \mathbb{N}$, but in this case both interpretations give the same value 1.)

Exercise 3.2. Let $n \in \mathbb{N}$.

(a) Prove that

$$(2n-1) \cdot (2n-3) \cdot \cdots \cdot 1 = \frac{(2n)!}{2^n n!}.$$

(The left hand side is understood to be the product of all odd integers from 1 to $2n-1$.)

(b) Prove that

$$\binom{-1/2}{n} = \left(\frac{-1}{4}\right)^n \binom{2n}{n}.$$

(c) Prove that

$$\binom{-1/3}{n}\binom{-2/3}{n} = \frac{(3n)!}{(3^n n!)^3}.$$

3.1.3. The recurrence relation of the binomial coefficients

Proposition 3.11. Let $m \in \mathbb{Q}$ and $n \in \{1,2,3,\ldots\}$. Then,

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n}. \tag{234}$$

Proof of Proposition 3.11. From $n \in \{1,2,3,\ldots\}$, we obtain $n! = n \cdot (n-1)!$, so that $(n-1)! = n!/n$ and thus $\dfrac{1}{(n-1)!} = \dfrac{1}{n!/n} = \dfrac{1}{n!} \cdot n$.

The definition of $\dbinom{m}{n-1}$ yields

$$\binom{m}{n-1} = \frac{m(m-1)\cdots(m-(n-1)+1)}{(n-1)!} = \frac{1}{(n-1)!} \cdot (m(m-1)\cdots(m-(n-1)+1)).$$

The same argument (applied to $m-1$ instead of m) yields

$$\begin{aligned}\binom{m-1}{n-1} &= \underbrace{\frac{1}{(n-1)!}}_{=\frac{1}{n!}\cdot n} \cdot \left((m-1)\underbrace{((m-1)-1)}_{=m-2}\cdots\underbrace{((m-1)-(n-1)+1)}_{=m-n+1}\right) \\ &= \frac{1}{n!} \cdot n \cdot ((m-1)(m-2)\cdots(m-n+1)).\end{aligned} \tag{235}$$

On the other hand,

$$\binom{m}{n} = \frac{m(m-1)\cdots(m-n+1)}{n!} = \frac{1}{n!}(m(m-1)\cdots(m-n+1)).$$

The same argument (applied to $m-1$ instead of m) yields

$$\begin{aligned}
\binom{m-1}{n} &= \frac{1}{n!}\left((m-1)\underbrace{((m-1)-1)}_{=m-2}\cdots\underbrace{((m-1)-n+1)}_{=m-n}\right) \\
&= \frac{1}{n!}\underbrace{((m-1)(m-2)\cdots(m-n))}_{=((m-1)(m-2)\cdots(m-n+1))\cdot(m-n)} \\
&= \frac{1}{n!}((m-1)(m-2)\cdots(m-n+1))\cdot(m-n) \\
&= \frac{1}{n!}(m-n)\cdot((m-1)(m-2)\cdots(m-n+1)).
\end{aligned}$$

Adding (235) to this equality, we obtain

$$\begin{aligned}
&\binom{m-1}{n}+\binom{m-1}{n-1} \\
&= \frac{1}{n!}(m-n)\cdot((m-1)(m-2)\cdots(m-n+1)) \\
&\qquad + \frac{1}{n!}\cdot n\cdot((m-1)(m-2)\cdots(m-n+1)) \\
&= \frac{1}{n!}\underbrace{((m-n)+n)}_{=m}\cdot((m-1)(m-2)\cdots(m-n+1)) \\
&= \frac{1}{n!}\underbrace{m\cdot((m-1)(m-2)\cdots(m-n+1))}_{=m(m-1)\cdots(m-n+1)} = \frac{1}{n!}(m(m-1)\cdots(m-n+1)) \\
&= \binom{m}{n} \qquad \left(\text{since } \binom{m}{n} = \frac{1}{n!}(m(m-1)\cdots(m-n+1))\right).
\end{aligned}$$

This proves Proposition 3.11. □

The formula (234) is known as the *recurrence relation of the binomial coefficients*[100].

Exercise 3.3. **(a)** Prove that every $n\in\mathbb{N}$ and $q\in\mathbb{Q}$ satisfy

$$\sum_{r=0}^{n}\binom{r+q}{r} = \binom{n+q+1}{n}.$$

(b) Prove that every $n\in\{-1,0,1,\ldots\}$ and $k\in\mathbb{N}$ satisfy

$$\sum_{i=0}^{n}\binom{i}{k} = \sum_{i=k}^{n}\binom{i}{k} = \binom{n+1}{k+1}.$$

[100] Often it is extended to the case $n=0$ by setting $\binom{m}{-1}=0$. It then follows from (227) in this case.

The formula (234) is responsible for the fact that "every number in Pascal's triangle is the sum of the two numbers above it". (Of course, if you use this fact as a *definition* of Pascal's triangle, then (234) is conversely responsible for the fact that the numbers in this triangle are the binomial coefficients.)

(Keep in mind that $\sum_{i=k}^{n} \binom{i}{k}$ is an empty sum whenever $n < k$.)

The claim of Exercise 3.3 **(b)** is one of several formulas known as the *hockey-stick identity* (due to the fact that marking the binomial coefficients appearing in it in Pascal's triangle results in a shape resembling a hockey stick[101]); it appears, e.g., in [Galvin17, Identity 11.10] (or, rather, the second equality sign of Exercise 3.3 **(b)** appears there, but the rest is easy).

3.1.4. The combinatorial interpretation of binomial coefficients

Proposition 3.12. If $m \in \mathbb{N}$ and $n \in \mathbb{N}$, and if S is an m-element set, then

$$\binom{m}{n} \text{ is the number of all } n\text{-element subsets of } S. \tag{236}$$

In less formal terms, Proposition 3.12 says the following: If $m \in \mathbb{N}$ and $n \in \mathbb{N}$, then $\binom{m}{n}$ is the number of ways to pick out n among m given objects, without replacement[102] and without regard for the order in which they are picked out. (Probabilists call this "unordered samples without replacement".)

Example 3.13. Proposition 3.12 (applied to $m = 4$, $n = 2$ and $S = \{0, 1, 2, 3\}$) shows that $\binom{4}{2}$ is the number of all 2-element subsets of $\{0, 1, 2, 3\}$ (since $\{0, 1, 2, 3\}$ is a 4-element set). And indeed, this is easy to verify by brute force: The 2-element subsets of $\{0, 1, 2, 3\}$ are

$$\{0, 1\}, \{0, 2\}, \{0, 3\}, \{1, 2\}, \{1, 3\} \text{ and } \{2, 3\},$$

so there are $6 = \binom{4}{2}$ of them.

Remark 3.14. Caution: Proposition 3.12 says nothing about binomial coefficients $\binom{m}{n}$ with negative m. Indeed, there are no m-element sets S when m is negative; thus, Proposition 3.12 would be vacuously true[103] when m is negative, but this would not help us computing binomial coefficients $\binom{m}{n}$ with negative m.

Actually, when $m \in \mathbb{Z}$ is negative, the number $\binom{m}{n}$ is positive for n even and negative for n odd (easy exercise), and so an interpretation of $\binom{m}{n}$ as a number

[101]See https://math.stackexchange.com/q/1490794 for an illustration.
[102]That is, one must not pick out the same object twice.

of ways to do something is rather unlikely. (On the other hand, $(-1)^n \binom{m}{n}$ does have such an interpretation.)

Remark 3.15. Some authors (for example, those of [LeLeMe16] and of [Galvin17]) use (236) as the *definition* of $\binom{m}{n}$. This is a legitimate definition of $\binom{m}{n}$ in the case when m and n are nonnegative integers (and, of course, equivalent to our definition); but it is not as general as ours, since it does not extend to negative (or non-integer) m.

Exercise 3.4. Prove Proposition 3.12.

Proposition 3.12 is one of the most basic facts of *enumerative combinatorics* – the part of mathematics that is mostly concerned with counting problems (i.e., the study of the sizes of finite sets). We will encounter some further results from enumerative combinatorics below (e.g., Exercise 3.15 and Exercise 4.3); but we shall not go deep into this subject. More serious expositions of enumerative combinatorics include Loehr's textbook [Loehr11], Galvin's lecture notes [Galvin17], Aigner's book [Aigner07], and Stanley's two-volume treatise ([Stanle11] and [Stanle01]).

3.1.5. Upper negation

Proposition 3.16. Let $m \in \mathbb{Q}$ and $n \in \mathbb{N}$. Then,

$$\binom{m}{n} = (-1)^n \binom{n-m-1}{n}. \tag{237}$$

[103]Recall that a mathematical statement of the form "if $\mathcal{A}$, then $\mathcal{B}$" is said to be *vacuously true* if $\mathcal{A}$ never holds. For example, the statement "if $0 = 1$, then every integer is odd" is vacuously true, because $0 = 1$ is false. Proposition 3.12 is vacuously true when m is negative, because the condition "S is an m-element set" never holds when m is negative.

By the laws of logic, a vacuously true statement is always true! See Convention 2.37 for a discussion of this principle.

Proof of Proposition 3.16. The equality (226) (applied to $n-m-1$ instead of m) yields

$$\begin{aligned}
\binom{n-m-1}{n} &= \frac{(n-m-1)((n-m-1)-1)\cdots((n-m-1)-n+1)}{n!} \\
&= \frac{1}{n!}(n-m-1)((n-m-1)-1)\cdots\underbrace{((n-m-1)-n+1)}_{=-m} \\
&= \frac{1}{n!}\underbrace{(n-m-1)((n-m-1)-1)\cdots(-m)}_{\substack{=(-m)(-m+1)\cdots(n-m-1) \\ \text{(here, we have just reversed the order of the factors in the product)}}} \\
&= \frac{1}{n!}\underbrace{(-m)}_{=(-1)m}\underbrace{(-m+1)}_{=(-1)(m-1)}\cdots\underbrace{(n-m-1)}_{=(-1)(m-n+1)} \\
&= \frac{1}{n!}((-1)m)((-1)(m-1))\cdots((-1)(m-n+1)) \\
&= \frac{1}{n!}(-1)^n(m(m-1)\cdots(m-n+1)),
\end{aligned}$$

so that

$$\begin{aligned}
(-1)^n\binom{n-m-1}{n} &= (-1)^n\cdot\frac{1}{n!}(-1)^n(m(m-1)\cdots(m-n+1)) \\
&= \underbrace{(-1)^n(-1)^n}_{\substack{=(-1)^{n+n}=(-1)^{2n}=1 \\ \text{(since } 2n \text{ is even)}}}\cdot\frac{1}{n!}(m(m-1)\cdots(m-n+1)) \\
&= \frac{1}{n!}(m(m-1)\cdots(m-n+1)) = \frac{m(m-1)\cdots(m-n+1)}{n!}.
\end{aligned}$$

Compared with (226), this yields $\binom{m}{n} = (-1)^n\binom{n-m-1}{n}$. Proposition 3.16 is therefore proven. □

The formula (237) is known as the *upper negation formula*.

Corollary 3.17. Let $n \in \mathbb{N}$. Then,

$$\binom{-1}{n} = (-1)^n.$$

Proof of Corollary 3.17. Proposition 3.16 (applied to $m = -1$) yields

$$\binom{-1}{n} = (-1)^n \binom{n-(-1)-1}{n} = (-1)^n \underbrace{\binom{n}{n}}_{\substack{=1 \\ \text{(by Proposition 3.9} \\ \text{(applied to } m=n\text{))}}} \qquad \left(\text{since } n-(-1)-1 = n\right)$$
$$= (-1)^n .$$

This proves Corollary 3.17. $\square$

Exercise 3.5. **(a)** Show that $\binom{-1}{k} = (-1)^k$ for each $k \in \mathbb{N}$.

(b) Show that $\binom{-2}{k} = (-1)^k (k+1)$ for each $k \in \mathbb{N}$.

(c) Show that $\dfrac{1! \cdot 2! \cdot \cdots \cdot (2n)!}{n!} = 2^n \cdot \left(\prod_{i=1}^{n} ((2i-1)!) \right)^2$ for each $n \in \mathbb{N}$.

Remark 3.18. Parts **(a)** and **(b)** of Exercise 3.5 are known facts (actually, part **(a)** is just a repetition of Corollary 3.17, for the purpose of making the analogy to part **(b)** more visible). Part **(c)** is a generalization of a puzzle posted on `https://www.reddit.com/r/math/comments/7rybhp/factorial_problem/` . (The puzzle boils down to the fact that $\dfrac{1! \cdot 2! \cdot \cdots \cdot (2n)!}{n!}$ is a perfect square when $n \in \mathbb{N}$ is even. But this follows from Exercise 3.5 **(c)**, because when $n \in \mathbb{N}$ is even, both factors 2^n and $\left(\prod_{i=1}^{n} ((2i-1)!) \right)^2$ on the right hand side of Exercise 3.5 **(c)** are perfect squares.)

3.1.6. Binomial coefficients of integers are integers

Lemma 3.19. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Then,

$$\binom{m}{n} \in \mathbb{N}.$$

Proof of Lemma 3.19. We have $m \in \mathbb{N}$. Thus, there exists an m-element set S (for example, $S = \{1, 2, \ldots, m\}$). Consider such an S. Then, $\binom{m}{n}$ is the number of all n-element subsets of S (because of (236)). Hence, $\binom{m}{n}$ is a nonnegative integer, so that $\binom{m}{n} \in \mathbb{N}$. This proves Lemma 3.19. $\square$

It is also easy to prove Lemma 3.19 by induction on m, using (227) and (231) in the induction base and using (234) in the induction step.

Proposition 3.20. Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. Then,

$$\binom{m}{n} \in \mathbb{Z}. \tag{238}$$

Proof of Proposition 3.20. We need to show (238). We are in one of the following two cases:

Case 1: We have $m \geq 0$.

Case 2: We have $m < 0$.

Let us first consider Case 1. In this case, we have $m \geq 0$. Hence, $m \in \mathbb{N}$. Thus, Lemma 3.19 yields $\binom{m}{n} \in \mathbb{N} \subseteq \mathbb{Z}$. This proves (238) in Case 1.

Let us now consider Case 2. In this case, we have $m < 0$. Thus, $m \leq -1$ (since m is an integer), so that $m+1 \leq 0$, so that $n-m-1 = n - \underbrace{(m+1)}_{\leq 0} \geq n \geq 0$. Hence, $n-m-1 \in \mathbb{N}$. Therefore, Lemma 3.19 (applied to $n-m-1$ instead of m) yields $\binom{n-m-1}{n} \in \mathbb{N} \subseteq \mathbb{Z}$. Now, (237) shows that $\binom{m}{n} = \underbrace{(-1)^n}_{\in \mathbb{Z}} \underbrace{\binom{n-m-1}{n}}_{\in \mathbb{Z}} \in \mathbb{Z}$ (here, we have used the fact that the product of two integers is an integer). This proves (238) in Case 2.

We thus have proven (238) in each of the two Cases 1 and 2. We can therefore conclude that (238) always holds. Thus, Proposition 3.20 is proven. □

The above proof of Proposition 3.20 may well be the simplest one. There is another proof, which uses Theorem 2.149, but it is more complicated[104]. There is yet another proof using basic number theory (specifically, checking how often a prime p appears in the numerator and the denominator of $\binom{m}{n} = \dfrac{m(m-1)\cdots(m-n+1)}{n!}$), but this is not quite easy.

3.1.7. The binomial formula

Proposition 3.21. Let x and y be two rational numbers (or real numbers, or complex numbers). Then,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{239}$$

for every $n \in \mathbb{N}$.

[104]It requires an induction on n nested inside the induction step of the induction on m.

Proposition 3.21 is the famous *binomial formula* (also known as the *binomial theorem*) and has a well-known standard proof by induction over n (using (234) and (227))[105]. Some versions of it hold for negative n as well (but not in the exact form (239), and not without restrictions).

Exercise 3.6. Prove Proposition 3.21.

There is an analogue of Proposition 3.21 for a sum of m rational numbers (rather than 2 rational numbers); it is called the "multinomial formula" (and involves the multinomial coefficients from Remark 3.10). We shall state it in a more general setting in Exercise 6.2.

3.1.8. The absorption identity

Proposition 3.22. Let $n \in \{1, 2, 3, \ldots\}$ and $m \in \mathbb{Q}$. Then,

$$\binom{m}{n} = \frac{m}{n} \binom{m-1}{n-1}. \tag{240}$$

Proof of Proposition 3.22. The definition of $\binom{m-1}{n-1}$ yields

$$\binom{m-1}{n-1} = \frac{(m-1)\left((m-1)-1\right)\cdots\left((m-1)-(n-1)+1\right)}{(n-1)!} = \frac{(m-1)(m-2)\cdots(m-n+1)}{(n-1)!}$$

(since $(m-1)-1 = m-2$ and $(m-1)-(n-1)+1 = m-n+1$). Multiplying both sides of this equality by $\frac{m}{n}$, we obtain

$$\frac{m}{n}\binom{m-1}{n-1} = \frac{m}{n} \cdot \frac{(m-1)(m-2)\cdots(m-n+1)}{(n-1)!} = \frac{m(m-1)(m-2)\cdots(m-n+1)}{n(n-1)!} = \frac{m(m-1)\cdots(m-n+1)}{n!}$$

(since $m(m-1)(m-2)\cdots(m-n+1) = m(m-1)\cdots(m-n+1)$ and $n(n-1)! = n!$). Compared with (226), this yields $\binom{m}{n} = \frac{m}{n}\binom{m-1}{n-1}$. This proves Proposition 3.22. □

The relation (240) is called the *absorption identity* in [GrKnPa94, §5.1].

[105] See Exercise 3.6 for this proof.

Exercise 3.7. Let k, a and b be three positive integers such that $k \leq a \leq b$. Prove that

$$\frac{k-1}{k} \sum_{n=a}^{b} \frac{1}{\binom{n}{k}} = \frac{1}{\binom{a-1}{k-1}} - \frac{1}{\binom{b}{k-1}}.$$

(In particular, all fractions appearing in this equality are well-defined.)

3.1.9. Trinomial revision

Proposition 3.23. Let $m \in \mathbb{Q}$, $a \in \mathbb{N}$ and $i \in \mathbb{N}$ be such that $i \geq a$. Then,

$$\binom{m}{i}\binom{i}{a} = \binom{m}{a}\binom{m-a}{i-a}. \tag{241}$$

Proof of Proposition 3.23. We have

$$\begin{aligned}
&\underbrace{\binom{m}{a}}_{\substack{=\frac{m(m-1)\cdots(m-a+1)}{a!} \\ \text{(by the definition of } \binom{m}{a}\text{)}}} \underbrace{\binom{m-a}{i-a}}_{\substack{=\frac{(m-a)((m-a)-1)\cdots((m-a)-(i-a)+1)}{(i-a)!} \\ \text{(by the definition of } \binom{m-a}{i-a}\text{)}}} \\
&= \frac{m(m-1)\cdots(m-a+1)}{a!} \cdot \frac{(m-a)((m-a)-1)\cdots((m-a)-(i-a)+1)}{(i-a)!} \\
&= \frac{1}{a! \cdot (i-a)!} \\
&\qquad \cdot \underbrace{(m(m-1)\cdots(m-a+1)) \cdot ((m-a)((m-a)-1)\cdots((m-a)-(i-a)+1))}_{\substack{=m(m-1)\cdots((m-a)-(i-a)+1) \\ =m(m-1)\cdots(m-i+1) \\ \text{(since } (m-a)-(i-a)=m-i\text{)}}} \\
&= \frac{1}{a! \cdot (i-a)!} \cdot m(m-1)\cdots(m-i+1).
\end{aligned}$$

Compared with

$$\underbrace{\binom{m}{i}}_{\substack{=\dfrac{m(m-1)\cdots(m-i+1)}{i!} \\ \text{(by the definition of } \binom{m}{i}\text{)}}} \underbrace{\binom{i}{a}}_{\substack{=\dfrac{i!}{a!\,(i-a)!} \\ \text{(by (229), applied to } i \text{ and } a \\ \text{instead of } m \text{ and } n\text{)}}}$$

$$= \frac{m(m-1)\cdots(m-i+1)}{i!}\cdot\frac{i!}{a!\,(i-a)!} = \frac{1}{a!\cdot(i-a)!}\cdot m(m-1)\cdots(m-i+1),$$

this yields $\binom{m}{i}\binom{i}{a} = \binom{m}{a}\binom{m-a}{i-a}$. This proves Proposition 3.23.

[Notice that we used (229) to simplify $\binom{i}{a}$ in this proof. Do not be tempted to use (229) to simplify $\binom{m}{i}$, $\binom{m}{a}$ and $\binom{m-a}{i-a}$: The m in these expressions may fail to be an integer!] □

Proposition 3.23 is a simple and yet highly useful formula, which Graham, Knuth and Patashnik call *trinomial revision* in [GrKnPa94, Table 174].

3.2. Binomial coefficients and polynomials

We have so far defined the binomial coefficient $\binom{m}{n}$ in the case when $n \in \mathbb{N}$ while m is some number (rational, real or complex). However, we can take this definition even further: For example, we can define $\binom{m}{n}$ when m is a polynomial with rational or real coefficients. Let us do this now:

Definition 3.24. Let $n \in \mathbb{N}$. Let m be a polynomial whose coefficients are rational numbers (or real numbers, or complex numbers).

We define a polynomial $\binom{m}{n}$ by the equality (226). This is a polynomial whose coefficients will be rational numbers or real numbers or complex numbers, depending on the coefficients of m.

Thus, in particular, for the polynomial $X \in \mathbb{Q}[X]$, we have

$$\binom{X}{n} = \frac{X(X-1)\cdots(X-n+1)}{n!} \qquad \text{for every } n \in \mathbb{N}.$$

In particular,

$$\binom{X}{0} = \frac{X(X-1)\cdots(X-0+1)}{0!} = \frac{(\text{empty product})}{1} = 1;$$
$$\binom{X}{1} = \frac{X(X-1)\cdots(X-1+1)}{1!} = \frac{X}{1} = X;$$
$$\binom{X}{2} = \frac{X(X-1)}{2!} = \frac{X(X-1)}{2} = \frac{1}{2}X^2 - \frac{1}{2}X;$$
$$\binom{X}{3} = \frac{X(X-1)(X-2)}{3!} = \frac{X(X-1)(X-2)}{6} = \frac{1}{6}X^3 - \frac{1}{2}X^2 + \frac{1}{3}X.$$

The polynomial $\binom{X}{n}$ lets us compute the binomial coefficients $\binom{m}{n}$ for all $m \in \mathbb{N}$, because of the following:

Proposition 3.25. Let $m \in \mathbb{Q}$ and $n \in \mathbb{N}$. Then, the rational number $\binom{m}{n}$ is the result of evaluating the polynomial $\binom{X}{n}$ at $X = m$.

Proof of Proposition 3.25. We have $\binom{X}{n} = \frac{X(X-1)\cdots(X-n+1)}{n!}$. Hence, the result of evaluating the polynomial $\binom{X}{n}$ at $X = m$ is

$$\frac{m(m-1)\cdots(m-n+1)}{n!} = \binom{m}{n} \qquad (\text{by (226)}).$$

This proves Proposition 3.25. □

We note the following properties of the polynomials $\binom{X}{n}$:

Proposition 3.26. (a) We have

$$\binom{X}{0} = 1.$$

(b) We have

$$\binom{X}{1} = X.$$

(c) For every $n \in \{1, 2, 3, \ldots\}$, we have

$$\binom{X}{n} = \binom{X-1}{n} + \binom{X-1}{n-1}.$$

(d) For every $n \in \mathbb{N}$, we have

$$\binom{X}{n} = (-1)^n \binom{n-X-1}{n}.$$

(e) For every $n \in \{1, 2, 3, \ldots\}$, we have

$$\binom{X}{n} = \frac{X}{n} \binom{X-1}{n-1}.$$

(f) Let $a \in \mathbb{N}$ and $i \in \mathbb{N}$ be such that $i \geq a$. Then,

$$\binom{X}{i} \binom{i}{a} = \binom{X}{a} \binom{X-a}{i-a}.$$

Proof of Proposition 3.26. **(a)** To obtain a proof of Proposition 3.26 **(a)**, replace every appearance of "m" by "X" in the proof of Proposition 3.3 **(a)**.

(b) To obtain a proof of Proposition 3.26 **(b)**, replace every appearance of "m" by "X" in the proof of Proposition 3.3 **(b)**.

(c) To obtain a proof of Proposition 3.26 **(c)**, replace every appearance of "m" by "X" in the proof of Proposition 3.11.

(d) To obtain a proof of Proposition 3.26 **(d)**, replace every appearance of "m" by "X" in the proof of Proposition 3.16.

(e) To obtain a proof of Proposition 3.26 **(e)**, replace every appearance of "m" by "X" in the proof of Proposition 3.22.

(f) To obtain a proof of Proposition 3.26 **(f)**, replace every appearance of "m" by "X" in the proof of Proposition 3.23. □

Recall that any polynomial $P \in \mathbb{Q}[X]$ (that is, any polynomial in the indeterminate X with rational coefficients) can be quasi-uniquely written in the form $P(X) = \sum_{i=0}^{d} c_i X^i$ with rational $c_0, c_1, \ldots, c_d$. The word "quasi-uniquely" here means that the coefficients $c_0, c_1, \ldots, c_d$ are uniquely determined when $d \in \mathbb{N}$ is specified; they are not literally unique because we can always increase d by adding new 0 coefficients (for example, the polynomial $(1+X)^2$ can be written both as $1 + 2X + X^2$ and as $1 + 2X + X^2 + 0X^3 + 0X^4$).

It is not hard to check that an analogue of this statement holds with the X^i replaced by the $\binom{X}{i}$:

Proposition 3.27. **(a)** Any polynomial $P \in \mathbb{Q}[X]$ can be quasi-uniquely written in the form $P(X) = \sum_{i=0}^{d} c_i \binom{X}{i}$ with rational $c_0, c_1, \ldots, c_d$. (Again, "quasi-uniquely" means that we can always increase d by adding new 0 coefficients, but apart from this the $c_0, c_1, \ldots, c_d$ are uniquely determined.)

(b) The polynomial P is *integer-valued* (i.e., its values at integers are integers) if and only if these rationals $c_0, c_1, \ldots, c_d$ are integers.

We will not use this fact below, but it gives context to Theorem 3.30 and Exercise 3.8 further below. The "if" part of Proposition 3.27 **(b)** follows from (238). For a full proof of Proposition 3.27 **(b)**, see [AndDos12, Theorem 10.3]. See also [daSilv12] for a proof of the "only if" part.

We shall now prove some facts and give some exercises about binomial coefficients; but let us first prove a fundamental property of polynomials:

Lemma 3.28. **(a)** Let P be a polynomial in the indeterminate X with rational coefficients. Assume that $P(x) = 0$ for all $x \in \mathbb{N}$. Then, $P = 0$ as polynomials[106].

(b) Let P and Q be two polynomials in the indeterminate X with rational coefficients. Assume that $P(x) = Q(x)$ for all $x \in \mathbb{N}$. Then, $P = Q$ as polynomials.

(c) Let P be a polynomial in the indeterminates X and Y with rational coefficients. Assume that $P(x, y) = 0$ for all $x \in \mathbb{N}$ and $y \in \mathbb{N}$. Then, $P = 0$ as polynomials.

(d) Let P and Q be two polynomials in the indeterminates X and Y with rational coefficients. Assume that $P(x, y) = Q(x, y)$ for all $x \in \mathbb{N}$ and $y \in \mathbb{N}$. Then, $P = Q$ as polynomials.

Lemma 3.28 is a well-known property of polynomials with rational coefficients; let us still prove it for the sake of completeness.

Proof of Lemma 3.28. **(a)** The polynomial P satisfies $P(x) = 0$ for every $x \in \mathbb{N}$. Hence, every $x \in \mathbb{N}$ is a root of P. Thus, the polynomial P has infinitely many roots. But a nonzero polynomial in one variable (with rational coefficients) can only have finitely many roots[107]. If P was nonzero, this would force a contradiction with the sentence before. So P must be zero. In other words, $P = 0$. Lemma 3.28 **(a)** is proven.

(b) Every $x \in \mathbb{N}$ satisfies $(P - Q)(x) = P(x) - Q(x) = 0$ (since $P(x) = Q(x)$). Hence, Lemma 3.28 **(a)** (applied to $P - Q$ instead of P) yields $P - Q = 0$. Thus, $P = Q$. Lemma 3.28 **(b)** is thus proven.

(c) Every $x \in \mathbb{N}$ and $y \in \mathbb{N}$ satisfy

$$P(x, y) = 0. \tag{242}$$

We can write the polynomial P in the form $P = \sum_{k=0}^{d} P_k(X) Y^k$, where d is an integer and where each $P_k(X)$ (for $0 \leq k \leq d$) is a polynomial in the single variable X. Consider this d and these $P_k(X)$.

[106] Recall that two polynomials are said to be equal if and only if their respective coefficients are equal.

[107] In fact, a stronger statement holds: A nonzero polynomial in one variable (with rational coefficients) having degree $n \geq 0$ has at most n roots. See, for example, [Goodma15, Corollary 1.8.24] or [Joyce17, Theorem 1.58] or [Walker87, Corollary 4.5.10] or [Elman18, Corollary 33.7] or [Swanso18, Theorem 2.4.13] or [Knapp16, Corollary 1.14] for a proof. Note that Swanson, in [Swanso18], works with polynomial functions instead of polynomials; but as far as roots are concerned, the difference does not matter (since the roots of a polynomial are precisely the roots of the corresponding polynomial function).

Fix $\alpha \in \mathbb{N}$. Every $x \in \mathbb{N}$ satisfies

$$P(\alpha, x) = \sum_{k=0}^{d} P_k(\alpha) x^k \qquad \left(\text{here, we substituted } \alpha \text{ and } x \text{ for } X \text{ and } Y \text{ in } P = \sum_{k=0}^{d} P_k(X) Y^k\right),$$

so that $\sum_{k=0}^{d} P_k(\alpha) x^k = P(\alpha, x) = 0$ (by (242), applied to α and x instead of x and y).

Therefore, Lemma 3.28 **(a)** (applied to $\sum_{k=0}^{d} P_k(\alpha) X^k$ instead of P) yields that

$$\sum_{k=0}^{d} P_k(\alpha) X^k = 0$$

as polynomials (in the indeterminate X). In other words, all coefficients of the polynomial $\sum_{k=0}^{d} P_k(\alpha) X^k$ are 0. In other words, $P_k(\alpha) = 0$ for all $k \in \{0, 1, \ldots, d\}$.

Now, let us forget that we fixed α. We thus have shown that $P_k(\alpha) = 0$ for all $k \in \{0, 1, \ldots, d\}$ and $\alpha \in \mathbb{N}$.

Let us now fix $k \in \{0, 1, \ldots, d\}$. Then, $P_k(\alpha) = 0$ for all $\alpha \in \mathbb{N}$. In other words, $P_k(x) = 0$ for all $x \in \mathbb{N}$. Hence, Lemma 3.28 **(a)** (applied to $P = P_k$) yields that $P_k = 0$ as polynomials.

Let us forget that we fixed k. We thus have proven that $P_k = 0$ as polynomials for each $k \in \{0, 1, \ldots, d\}$. Hence, $P = \sum_{k=0}^{d} \underbrace{P_k(X)}_{=0} Y^k = 0$. This proves Lemma 3.28 **(c)**.

(d) Every $x \in \mathbb{N}$ and $y \in \mathbb{N}$ satisfy

$$(P - Q)(x, y) = P(x, y) - Q(x, y) = 0 \qquad (\text{since } P(x, y) = Q(x, y)).$$

Hence, Lemma 3.28 **(c)** (applied to $P - Q$ instead of P) yields $P - Q = 0$. Thus, $P = Q$. Lemma 3.28 **(d)** is proven. □

Of course, Lemma 3.28 can be generalized to polynomials in more than two variables (the proof of Lemma 3.28 **(c)** essentially suggests how to prove this generalization by induction over the number of variables).[108]

[108] If you know what a commutative ring is, you might wonder whether Lemma 3.28 can also be generalized to polynomials with coefficients from other commutative rings (e.g., from $\mathbb{R}$ or $\mathbb{C}$) instead of rational coefficients. In other words, what happens if we replace "rational coefficients" by "coefficients in R" throughout Lemma 3.28, where R is some commutative ring? (Of course, we will then have to also replace $P(x)$ by $P(x \cdot 1_R)$ and so on.)

The answer is that Lemma 3.28 becomes generally false if we don't require anything more specific on R. However, there are certain conditions on R that make Lemma 3.28 remain valid.

3.3. The Chu-Vandermonde identity

3.3.1. The statements

The following fact is known as the *Chu-Vandermonde identity*[109]:

Theorem 3.29. Let $n \in \mathbb{N}$, $x \in \mathbb{Q}$ and $y \in \mathbb{Q}$. Then,

$$\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}.$$

Let us also give an analogous statement for polynomials:

Theorem 3.30. Let $n \in \mathbb{N}$. Then,

$$\binom{X+Y}{n} = \sum_{k=0}^{n} \binom{X}{k} \binom{Y}{n-k}$$

(an equality between polynomials in two variables X and Y).

We will give two proofs of this theorem: one algebraic, and one combinatorial.[110]

3.3.2. An algebraic proof

First proof of Theorem 3.29. Forget that we fixed n, x and y. We thus must prove that for each $n \in \mathbb{N}$, we have

$$\left(\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k} \text{ for all } x \in \mathbb{Q} \text{ and } y \in \mathbb{Q} \right). \qquad (243)$$

We shall prove (243) by induction over n:

For instance, Lemma 3.28 remains valid for $R = \mathbb{Z}$, for $R = \mathbb{R}$ and for $R = \mathbb{C}$, as well as for R being any polynomial ring over $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. More generally, Lemma 3.28 is valid if R is any field of characteristic 0 (i.e., any field such that the elements $n \cdot 1_R$ for n ranging over $\mathbb{N}$ are pairwise distinct), or any subring of such a field.

[109]See the Wikipedia page for part of its history. Usually, the equality $\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}$ for two **nonnegative integers** x and y (this is a particular case of Theorem 3.29) is called the *Vandermonde identity* (or the *Vandermonde convolution identity*), whereas the name "*Chu-Vandermonde identity*" is used for the identity $\binom{X+Y}{n} = \sum_{k=0}^{n} \binom{X}{k} \binom{Y}{n-k}$ in which X and Y are **indeterminates** (this is Theorem 3.30). However, this seems to be mostly a matter of convention (which isn't even universally followed); and anyway the two identities are easily derived from one another as we will see in the second proof of Theorem 3.30.

[110]Note that Theorem 3.29 appears in [GrKnPa94, (5.27)], where it is called *Vandermonde's convolution*. The second proof of Theorem 3.29 we shall show below is just a more detailed writeup of the proof given there.

Induction base: Let $x \in \mathbb{Q}$ and $y \in \mathbb{Q}$. Proposition 3.3 **(a)** (applied to $m = x$) yields $\dbinom{x}{0} = 1$. Proposition 3.3 **(a)** (applied to $m = y$) yields $\dbinom{y}{0} = 1$. Now,

$$\sum_{k=0}^{0} \binom{x}{k} \binom{y}{0-k} = \underbrace{\binom{x}{0}}_{=1} \underbrace{\binom{y}{0-0}}_{=\binom{y}{0}=1} = 1. \qquad (244)$$

But Proposition 3.3 **(a)** (applied to $m = x + y$) yields $\dbinom{x+y}{0} = 1$. Compared with (244), this yields $\dbinom{x+y}{0} = \sum\limits_{k=0}^{0} \dbinom{x}{k} \dbinom{y}{0-k}$.

Now, forget that we fixed x and y. We thus have shown that

$$\binom{x+y}{0} = \sum_{k=0}^{0} \binom{x}{k} \binom{y}{0-k} \qquad \text{for all } x \in \mathbb{Q} \text{ and } y \in \mathbb{Q}.$$

In other words, (243) holds for $n = 0$. This completes the induction base.

Induction step: Let N be a positive integer. Assume that (243) holds for $n = N - 1$. We need to prove that (243) holds for $n = N$. In other words, we need to prove that

$$\binom{x+y}{N} = \sum_{k=0}^{N} \binom{x}{k} \binom{y}{N-k} \qquad \text{for all } x \in \mathbb{Q} \text{ and } y \in \mathbb{Q}. \qquad (245)$$

We have assumed that (243) holds for $n = N - 1$. In other words, we have

$$\binom{x+y}{N-1} = \sum_{k=0}^{N-1} \binom{x}{k} \binom{y}{(N-1)-k} \qquad \text{for all } x \in \mathbb{Q} \text{ and } y \in \mathbb{Q}. \qquad (246)$$

Now, let us prove (245):

[*Proof of* (245): Fix $x \in \mathbb{Q}$ and $y \in \mathbb{Q}$. Then, (246) (applied to $x - 1$ instead of x) yields

$$\begin{aligned}
\binom{x-1+y}{N-1} &= \sum_{k=0}^{N-1} \binom{x-1}{k} \binom{y}{(N-1)-k} \\
&= \sum_{k=1}^{N} \binom{x-1}{k-1} \underbrace{\binom{y}{(N-1)-(k-1)}}_{=\binom{y}{N-k}} \\
&\qquad \text{(here, we have substituted } k-1 \text{ for } k \text{ in the sum)} \\
&= \sum_{k=1}^{N} \binom{x-1}{k-1} \binom{y}{N-k}.
\end{aligned}$$

Since $x-1+y=x+y-1$, this rewrites as

$$\binom{x+y-1}{N-1}=\sum_{k=1}^{N}\binom{x-1}{k-1}\binom{y}{N-k}. \tag{247}$$

On the other hand, (246) (applied to $y-1$ instead of y) shows that

$$\binom{x+y-1}{N-1}=\sum_{k=0}^{N-1}\binom{x}{k}\underbrace{\binom{y-1}{(N-1)-k}}_{=\binom{y-1}{N-k-1}}=\sum_{k=0}^{N-1}\binom{x}{k}\binom{y-1}{N-k-1}. \tag{248}$$

Next, we notice a simple consequence of (240): We have

$$\frac{x}{N}\binom{x-1}{a-1}=\frac{a}{N}\binom{x}{a} \qquad \text{for every } a\in\{1,2,3,\ldots\} \tag{249}$$

[111]. The same argument (applied to y instead of x) shows that

$$\frac{y}{N}\binom{y-1}{a-1}=\frac{a}{N}\binom{y}{a} \qquad \text{for every } a\in\{1,2,3,\ldots\}. \tag{250}$$

We have

$$\begin{aligned}\frac{x}{N}\underbrace{\binom{x+y-1}{N-1}}_{\substack{=\sum_{k=1}^{N}\binom{x-1}{k-1}\binom{y}{N-k}\\ \text{(by (247))}}} &= \frac{x}{N}\sum_{k=1}^{N}\binom{x-1}{k-1}\binom{y}{N-k}\\ &=\sum_{k=1}^{N}\underbrace{\frac{x}{N}\binom{x-1}{k-1}}_{\substack{=\frac{k}{N}\binom{x}{k}\\ \text{(by (249),}\\ \text{applied to } a=k)}}\binom{y}{N-k}=\sum_{k=1}^{N}\frac{k}{N}\binom{x}{k}\binom{y}{N-k}.\end{aligned}$$

[111] *Proof of (249):* Let $a\in\{1,2,3,\ldots\}$. Then, (240) (applied to $m=x$ and $n=a$) yields $\binom{x}{a}=\frac{x}{a}\binom{x-1}{a-1}$. Hence,

$$\frac{a}{N}\underbrace{\binom{x}{a}}_{=\frac{x}{a}\binom{x-1}{a-1}}=\underbrace{\frac{a}{N}\cdot\frac{x}{a}}_{=\frac{x}{N}}\binom{x-1}{a-1}=\frac{x}{N}\binom{x-1}{a-1}.$$

This proves (249).

Compared with

$$\sum_{k=0}^{N} \frac{k}{N}\binom{x}{k}\binom{y}{N-k} = \underbrace{\frac{0}{N}\binom{x}{0}\binom{y}{N-0}}_{=0} + \sum_{k=1}^{N} \frac{k}{N}\binom{x}{k}\binom{y}{N-k}$$

(here, we have split off the addend for $k=0$)

$$= \sum_{k=1}^{N} \frac{k}{N}\binom{x}{k}\binom{y}{N-k},$$

this yields

$$\frac{x}{N}\binom{x+y-1}{N-1} = \sum_{k=0}^{N} \frac{k}{N}\binom{x}{k}\binom{y}{N-k}. \tag{251}$$

We also have

$$\frac{y}{N}\underbrace{\binom{x+y-1}{N-1}}_{\substack{=\sum_{k=0}^{N-1}\binom{x}{k}\binom{y-1}{N-k-1} \\ \text{(by (248))}}} = \frac{y}{N}\sum_{k=0}^{N-1}\binom{x}{k}\binom{y-1}{N-k-1}$$

$$= \sum_{k=0}^{N-1}\binom{x}{k}\underbrace{\frac{y}{N}\binom{y-1}{N-k-1}}_{\substack{=\frac{N-k}{N}\binom{y}{N-k} \\ \text{(by (250), applied to } a=N-k \\ \text{(since } N-k\in\{1,2,3,\ldots\} \text{ (because } k\in\{0,1,\ldots,N-1\})))}}$$

$$= \sum_{k=0}^{N-1}\binom{x}{k}\frac{N-k}{N}\binom{y}{N-k} = \sum_{k=0}^{N-1}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k}.$$

Compared with

$$\sum_{k=0}^{N}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k} = \sum_{k=0}^{N-1}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k} + \underbrace{\frac{N-N}{N}}_{=0}\binom{x}{N}\binom{y}{N-N}$$

(here, we have split off the addend for $k=N$)

$$= \sum_{k=0}^{N-1}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k} + \underbrace{0\binom{x}{N}\binom{y}{N-N}}_{=0}$$

$$= \sum_{k=0}^{N-1}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k},$$

this yields

$$\frac{y}{N}\binom{x+y-1}{N-1} = \sum_{k=0}^{N} \frac{N-k}{N}\binom{x}{k}\binom{y}{N-k}. \tag{252}$$

Now, (240) (applied to $m = x + y$ and $n = N$) yields

$$\begin{aligned}
\binom{x+y}{N} &= \underbrace{\frac{x+y}{N}}_{=\frac{x}{N}+\frac{y}{N}} \binom{x+y-1}{N-1} = \left(\frac{x}{N}+\frac{y}{N}\right)\binom{x+y-1}{N-1} \\
&= \underbrace{\frac{x}{N}\binom{x+y-1}{N-1}}_{\substack{=\sum_{k=0}^{N}\frac{k}{N}\binom{x}{k}\binom{y}{N-k} \\ \text{(by (251))}}} + \underbrace{\frac{y}{N}\binom{x+y-1}{N-1}}_{\substack{=\sum_{k=0}^{N}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k} \\ \text{(by (252))}}} \\
&= \sum_{k=0}^{N}\frac{k}{N}\binom{x}{k}\binom{y}{N-k} + \sum_{k=0}^{N}\frac{N-k}{N}\binom{x}{k}\binom{y}{N-k} \\
&= \sum_{k=0}^{N}\left(\underbrace{\frac{k}{N}+\frac{N-k}{N}}_{=1}\right)\binom{x}{k}\binom{y}{N-k} = \sum_{k=0}^{N}\binom{x}{k}\binom{y}{N-k}.
\end{aligned}$$

This proves (245).]

We have thus proven (245). In other words, (243) holds for $n = N$. This completes the induction step. Thus, the induction proof of (243) is complete. Hence, Theorem 3.29 is proven. □

The above proof has the advantage of being completely algebraic; it thus does not rely on what x and y actually are. It works equally well if x and y are assumed to be real numbers or complex numbers or polynomials. Thus, it can also be used to prove Theorem 3.30:

First proof of Theorem 3.30. To obtain a proof of Theorem 3.30, replace every appearance of "x" by "X" and every appearance of "y" by "Y" in the above First proof of Theorem 3.29. □

3.3.3. A combinatorial proof

We shall next give a different, combinatorial proof of Theorems 3.29 and 3.30. This proof is somewhat indirect, as it begins by showing the following particular case of Theorem 3.29:

Lemma 3.31. Let $n \in \mathbb{N}$, $x \in \mathbb{N}$ and $y \in \mathbb{N}$. Then,

$$\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}. \tag{253}$$

This lemma is less general than Theorem 3.29, since it requires x and y to belong to $\mathbb{N}$.

Proof of Lemma 3.31. For every $N \in \mathbb{N}$, we let $[N]$ denote the N-element set $\{1, 2, \ldots, N\}$.

Recall that $\binom{x+y}{n}$ is the number of n-element subsets of a given $(x+y)$-element set[112]. Since $[x+y]$ is an $(x+y)$-element set, we thus conclude that $\binom{x+y}{n}$ is the number of n-element subsets of $[x+y]$.

But let us count the n-element subsets of $[x+y]$ in a different way (i.e., find a different expression for the number of n-element subsets of $[x+y]$). Namely, we can choose an n-element subset S of $[x+y]$ by means of the following process:

1. We decide how many elements of this subset S will be among the numbers $1, 2, \ldots, x$. Let k be the number of these elements. Clearly, k must be an integer between 0 and n (inclusive)[113].

2. Then, we choose these k elements of S among the numbers $1, 2, \ldots, x$. This can be done in $\binom{x}{k}$ different ways (because we are choosing k out of x numbers, with no repetitions, and with no regard for their order; in other words, we are choosing a k-element subset of $\{1, 2, \ldots, x\}$).

3. Then, we choose the remaining $n-k$ elements of S (because S should have n elements in total) among the remaining numbers $x+1, x+2, \ldots, x+y$. This can be done in $\binom{y}{n-k}$ ways (because we are choosing $n-k$ out of y numbers, with no repetitions, and with no regard for their order).

This process makes it clear that the total number of ways to choose an n-element subset S of $[x+y]$ is $\sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}$. In other words, the number of n-element subsets of $[x+y]$ is $\sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}$. But earlier, we have shown that the same

[112]This follows from (236).

[113]Because the subset S will have n elements in total, and thus at most n of them can be among the numbers $1, 2, \ldots, x$.

number is $\dbinom{x+y}{n}$. Comparing these two results, we conclude that $\dbinom{x+y}{n} = \sum\limits_{k=0}^{n} \dbinom{x}{k} \dbinom{y}{n-k}$. Thus, Lemma 3.31 is proven. $\square$

We now shall leverage Lemma 3.28 to derive Theorem 3.30 from this lemma:

Second proof of Theorem 3.30. We define two polynomials P and Q in the indeterminates X and Y with rational coefficients by setting

$$P = \dbinom{X+Y}{n};$$
$$Q = \sum_{k=0}^{n} \dbinom{X}{k} \dbinom{Y}{n-k}$$

[114]. The equality (253) (which we have proven) states that $P(x,y) = Q(x,y)$ for all $x \in \mathbb{N}$ and $y \in \mathbb{N}$. Thus, Lemma 3.28 **(d)** yields that $P = Q$. Recalling how P and Q are defined, we can rewrite this as $\dbinom{X+Y}{n} = \sum\limits_{k=0}^{n} \dbinom{X}{k} \dbinom{Y}{n-k}$. This proves Theorem 3.30. $\square$

The argument that we used at the end of the above proof to derive Theorem 3.30 from (253) is a very common argument that appears in proofs of equalities for binomial coefficients. The binomial coefficients $\dbinom{m}{n}$ are defined for arbitrary rational, real or complex m [115], but their combinatorial interpretation (via counting subsets) only makes sense when m and n are nonnegative integers. Thus, if we want to prove an identity of the form $P = Q$ (where P and Q are two polynomials, say, in two indeterminates X and Y) using the combinatorial interpretation of binomial coefficients, then a reasonable tactic is to first show that $P(x,y) = Q(x,y)$ for all $x \in \mathbb{N}$ and $y \in \mathbb{N}$ (using combinatorics), and then to use something like Lemma 3.28 in order to conclude that P and Q are equal as polynomials. We shall see this tactic used a few more times.[116]

[114] These are both polynomials since $\dbinom{X+Y}{n}$, $\dbinom{X}{k}$ and $\dbinom{Y}{n-k}$ are polynomials in X and Y.

[115] For example, terms like $\dbinom{-1/2}{3}$, $\dbinom{2+\sqrt{3}}{5}$ and $\dbinom{-7}{0}$ make perfect sense. (But we cannot substitute arbitrary complex numbers for n in $\dbinom{m}{n}$. So far we have only defined $\dbinom{m}{n}$ for $n \in \mathbb{N}$. It is usual to define $\dbinom{m}{n}$ to mean 0 for negative integers n, and using analysis (specifically, the Γ function) it is possible to give a reasonable meaning to $\dbinom{m}{n}$ for m and n being reals, but this will no longer be a polynomial in m.)

[116] This tactic is called "the polynomial argument" in [GrKnPa94, §5.1].

Second proof of Theorem 3.29. Theorem 3.30 yields

$$\binom{X+Y}{n} = \sum_{k=0}^{n} \binom{X}{k} \binom{Y}{n-k}$$

(an equality between polynomials in two variables X and Y). Now, let us evaluate both sides of this equality at $X = x$ and $Y = y$. As a result, we obtain

$$\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}$$

(because of Proposition 3.25). □

3.3.4. Some applications

Let us give some sample applications of Theorem 3.29:

Proposition 3.32. (a) For every $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ and $n \in \mathbb{N}$, we have

$$\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}.$$

(b) For every $x \in \mathbb{N}$ and $y \in \mathbb{Z}$, we have

$$\binom{x+y}{x} = \sum_{k=0}^{x} \binom{x}{k} \binom{y}{k}.$$

(c) For every $n \in \mathbb{N}$, we have

$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2.$$

(d) For every $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ and $n \in \mathbb{N}$, we have

$$\binom{x-y}{n} = \sum_{k=0}^{n} (-1)^k \binom{x}{n-k} \binom{k+y-1}{k}.$$

(e) For every $x \in \mathbb{N}$ and $y \in \mathbb{Z}$ and $n \in \mathbb{N}$ with $x \leq n$, we have

$$\binom{y-x-1}{n-x} = \sum_{k=0}^{n} (-1)^{k-x} \binom{k}{x} \binom{y}{n-k}.$$

(f) For every $x \in \mathbb{N}$ and $y \in \mathbb{N}$ and $n \in \mathbb{N}$, we have

$$\binom{n+1}{x+y+1} = \sum_{k=0}^{n} \binom{k}{x} \binom{n-k}{y}.$$

(g) For every $x \in \mathbb{Z}$ and $y \in \mathbb{N}$ and $n \in \mathbb{N}$ satisfying $x + y \geq 0$ and $n \geq x$, we have

$$\binom{x+y}{n} = \sum_{k=0}^{x+y} \binom{x}{k} \binom{y}{n+k-x}.$$

Remark 3.33. I have learnt Proposition 3.32 **(f)** from the AoPS forum. Proposition 3.32 **(g)** is a generalization of Proposition 3.32 **(b)**.

Note that if we apply Proposition 3.32 **(f)** to $y = 0$, then we obtain the identity $\dbinom{n+1}{x+1} = \sum_{k=0}^{n} \dbinom{k}{x}$ for all $n \in \mathbb{N}$ and $x \in \mathbb{N}$. This identity is also a particular case of Exercise 3.3 **(b)**.

Proof of Proposition 3.32. **(a)** Proposition 3.32 **(a)** is a particular case of Theorem 3.29.

(b) Let $x \in \mathbb{N}$ and $y \in \mathbb{Z}$. Proposition 3.32 **(a)** (applied to y, x and x instead of x, y and n) yields

$$\binom{y+x}{x} = \sum_{k=0}^{x} \binom{y}{k} \binom{x}{x-k}.$$

Compared with

$$\sum_{k=0}^{x} \underbrace{\binom{x}{k}}_{\substack{=\binom{x}{x-k} \\ \text{(by (232), applied to } m=x \text{ and } n=k \\ \text{(since } x \geq k \text{ (because } k \leq x)))}} \binom{y}{k} = \sum_{k=0}^{x} \binom{x}{x-k} \binom{y}{k} = \sum_{k=0}^{x} \binom{y}{k} \binom{x}{x-k},$$

this yields $\dbinom{y+x}{x} = \sum_{k=0}^{x} \dbinom{x}{k} \dbinom{y}{k}$. Since $y + x = x + y$, this rewrites as $\dbinom{x+y}{x} = \sum_{k=0}^{x} \dbinom{x}{k} \dbinom{y}{k}$. This proves Proposition 3.32 **(b)**.

(c) Let $n \in \mathbb{N}$. Applying Proposition 3.32 **(b)** to $x = n$ and $y = n$, we obtain

$$\binom{n+n}{n} = \sum_{k=0}^{n} \underbrace{\binom{n}{k} \binom{n}{k}}_{=\binom{n}{k}^2} = \sum_{k=0}^{n} \binom{n}{k}^2.$$

Since $n + n = 2n$, this rewrites as $\dbinom{2n}{n} = \sum_{k=0}^{n} \dbinom{n}{k}^2$. This proves Proposition 3.32 **(c)**.

(d) Let $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ and $n \in \mathbb{N}$. Proposition 3.32 **(a)** (applied to $-y$ instead of y) yields

$$\begin{aligned}
\binom{x+(-y)}{n} &= \sum_{k=0}^{n} \binom{x}{k}\binom{-y}{n-k} = \sum_{k=0}^{n} \binom{x}{n-k} \underbrace{\binom{-y}{n-(n-k)}}_{=\binom{-y}{k}} \\
&\qquad \text{(here, we substituted } n-k \text{ for } k \text{ in the sum)} \\
&= \sum_{k=0}^{n} \binom{x}{n-k} \underbrace{\binom{-y}{k}}_{\substack{=(-1)^k \binom{k-(-y)-1}{k} \\ \text{(by (237), applied to } k \text{ and } -y \text{ instead of } n \text{ and } m\text{)}}} \\
&= \sum_{k=0}^{n} \underbrace{\binom{x}{n-k}(-1)^k}_{=(-1)^k\binom{x}{n-k}} \underbrace{\binom{k-(-y)-1}{k}}_{=\binom{k+y-1}{k}} \\
&= \sum_{k=0}^{n} (-1)^k \binom{x}{n-k}\binom{k+y-1}{k}.
\end{aligned}$$

Since $x+(-y)=x-y$, this rewrites as $\binom{x-y}{n} = \sum_{k=0}^{n} (-1)^k \binom{x}{n-k}\binom{k+y-1}{k}$. This proves Proposition 3.32 **(d)**.

(e) Let $x \in \mathbb{N}$ and $y \in \mathbb{Z}$ and $n \in \mathbb{N}$ be such that $x \leq n$. From $x \in \mathbb{N}$, we obtain $0 \leq x$ and thus $0 \leq x \leq n$. We notice that every integer $k \geq x$ satisfies

$$\binom{k}{k-x} = \binom{k}{x} \tag{254}$$

[117]. Furthermore, $n - x \in \mathbb{N}$ (since $x \leq n$). Hence, we can apply Proposition 3.32

[117] *Proof of (254):* Let k be an integer such that $k \geq x$. Thus, $k - x \in \mathbb{N}$. Also, $k \geq x \geq 0$ (since $x \in \mathbb{N}$), and thus $k \in \mathbb{N}$. Now, recall that $k \geq x$. Hence, (232) (applied to k and x instead of m and n) yields $\binom{k}{x} = \binom{k}{k-x}$. This proves (254).

(a) to y, $-x-1$ and $n-x$ instead of x, y and n. As a result, we obtain

$$\begin{aligned}
\dbinom{y+(-x-1)}{n-x} &= \sum_{k=0}^{n-x} \dbinom{y}{k} \underbrace{\dbinom{-x-1}{(n-x)-k}}_{\substack{=(-1)^{(n-x)-k}\dbinom{((n-x)-k)-(-x-1)-1}{(n-x)-k} \\ \text{(by (237), applied to } -x-1 \text{ and } (n-x)-k \\ \text{instead of } m \text{ and } n)}} \\
&= \sum_{k=0}^{n-x} \dbinom{y}{k} (-1)^{(n-x)-k} \underbrace{\dbinom{((n-x)-k)-(-x-1)-1}{(n-x)-k}}_{\substack{=\dbinom{n-k}{(n-x)-k} \\ \text{(since } ((n-x)-k)-(-x-1)-1=n-k)}} \\
&= \sum_{k=0}^{n-x} \dbinom{y}{k} (-1)^{(n-x)-k} \dbinom{n-k}{(n-x)-k} \\
&= \underbrace{\sum_{k=n-(n-x)}^{n}}_{\substack{=\sum_{k=x}^{n} \\ \text{(since } n-(n-x)=x)}} \dbinom{y}{n-k} \underbrace{(-1)^{(n-x)-(n-k)}}_{\substack{=(-1)^{k-x} \\ \text{(since } (n-x)-(n-k)=k-x)}} \underbrace{\dbinom{n-(n-k)}{(n-x)-(n-k)}}_{\substack{=\dbinom{k}{k-x} \\ \text{(since } n-(n-k)=k \\ \text{and } (n-x)-(n-k)=k-x)}} \\
&\qquad \text{(here, we have substituted } n-k \text{ for } k \text{ in the sum)} \\
&= \sum_{k=x}^{n} \dbinom{y}{n-k} (-1)^{k-x} \underbrace{\dbinom{k}{k-x}}_{\substack{=\dbinom{k}{x} \\ \text{(by (254))}}} = \sum_{k=x}^{n} \dbinom{y}{n-k} (-1)^{k-x} \dbinom{k}{x} \\
&= \sum_{k=x}^{n} (-1)^{k-x} \dbinom{k}{x} \dbinom{y}{n-k}.
\end{aligned}$$

Compared with

$$\begin{aligned}
&\sum_{k=0}^{n}(-1)^{k-x}\binom{k}{x}\binom{y}{n-k} \\
&= \sum_{k=0}^{x-1}(-1)^{k-x}\underbrace{\binom{k}{x}}_{\substack{=0 \\ \text{(by (231), applied to } k \text{ and } x \\ \text{instead of } m \text{ and } n \text{ (since } k\leq x-1<x))}}\binom{y}{n-k}+\sum_{k=x}^{n}(-1)^{k-x}\binom{k}{x}\binom{y}{n-k} \\
&\qquad (\text{since } 0\leq x\leq n) \\
&= \underbrace{\sum_{k=0}^{x-1}(-1)^{k-x}0\binom{y}{n-k}}_{=0}+\sum_{k=x}^{n}(-1)^{k-x}\binom{k}{x}\binom{y}{n-k}=\sum_{k=x}^{n}(-1)^{k-x}\binom{k}{x}\binom{y}{n-k},
\end{aligned}$$

this yields

$$\binom{y+(-x-1)}{n-x}=\sum_{k=0}^{n}(-1)^{k-x}\binom{k}{x}\binom{y}{n-k}.$$

In other words,

$$\binom{y-x-1}{n-x}=\sum_{k=0}^{n}(-1)^{k-x}\binom{k}{x}\binom{y}{n-k}$$

(since $y+(-x-1)=y-x-1$). This proves Proposition 3.32 **(e)**.

(f) Let $x\in\mathbb{N}$ and $y\in\mathbb{N}$ and $n\in\mathbb{N}$. We must be in one of the following two cases:

Case 1: We have $n<x+y$.

Case 2: We have $n\geq x+y$.

Let us first consider Case 1. In this case, we have $n<x+y$. Thus, $n+1<x+y+1$. Therefore, $\binom{n+1}{x+y+1}=0$ (by (231), applied to $n+1$ and $x+y+1$ instead of m and n). But every $k\in\{0,1,\ldots,n\}$ satisfies $\binom{k}{x}\binom{n-k}{y}=0$ [118].

---

[118] *Proof.* Let $k\in\{0,1,\ldots,n\}$. We need to show that $\binom{k}{x}\binom{n-k}{y}=0$.

If we have $k<x$, then we have $\binom{k}{x}=0$ (by (231), applied to k and x instead of m and n). Therefore, if we have $k<x$, then $\underbrace{\binom{k}{x}}_{=0}\binom{n-k}{y}=0$. Hence, for the rest of this proof of $\binom{k}{x}\binom{n-k}{y}=0$, we can WLOG assume that we don't have $k<x$. Assume this.

We have $k\leq n$ (since $k\in\{0,1,\ldots,n\}$) and thus $n-k\in\mathbb{N}$.

We have $k\geq x$ (since we don't have $k<x$), and thus $n-\underbrace{k}_{\geq x}\leq n-x<y$ (since $n<$

Hence, $\sum_{k=0}^{n} \underbrace{\dbinom{k}{x}\dbinom{n-k}{y}}_{=0} = \sum_{k=0}^{n} 0 = 0$. Compared with $\dbinom{n+1}{x+y+1} = 0$, this yields $\dbinom{n+1}{x+y+1} = \sum_{k=0}^{n} \dbinom{k}{x}\dbinom{n-k}{y}$. Thus, Proposition 3.32 **(f)** is proven in Case 1.

Let us now consider Case 2. In this case, we have $n \geq x+y$. Hence, $n-y \geq x$ (since $x \in \mathbb{N}$), so that $(n-y)-x \in \mathbb{N}$. Also, $n-y \geq x \geq 0$ and thus $n-y \in \mathbb{N}$. Moreover, $x \leq n-y$. Therefore, we can apply Proposition 3.32 **(e)** to $-y-1$ and $n-y$ instead of y and n. As a result, we obtain

$$\begin{aligned}
&\dbinom{(-y-1)-x-1}{(n-y)-x} \\
&= \sum_{k=0}^{n-y} (-1)^{k-x} \dbinom{k}{x} \underbrace{\dbinom{-y-1}{(n-y)-k}}_{\substack{=(-1)^{(n-y)-k}\dbinom{((n-y)-k)-(-y-1)-1}{(n-y)-k} \\ \text{(by (237), applied to } -y-1 \text{ and } (n-y)-k \\ \text{instead of } m \text{ and } n)}} \\
&= \sum_{k=0}^{n-y} \underbrace{(-1)^{k-x}\dbinom{k}{x}(-1)^{(n-y)-k}}_{=(-1)^{(n-y)-k}\dbinom{k}{x}} \underbrace{\dbinom{((n-y)-k)-(-y-1)-1}{(n-y)-k}}_{\substack{=\dbinom{n-k}{(n-y)-k} \\ \text{(since } ((n-y)-k)-(-y-1)-1=n-k)}} \\
&= \sum_{k=0}^{n-y} \underbrace{(-1)^{k-x}(-1)^{(n-y)-k}}_{\substack{=(-1)^{(k-x)+((n-y)-k)}=(-1)^{n-x-y} \\ \text{(since } (k-x)+((n-y)-k)=n-x-y)}} \dbinom{k}{x}\dbinom{n-k}{(n-y)-k} \\
&= \sum_{k=0}^{n-y} (-1)^{n-x-y} \dbinom{k}{x}\dbinom{n-k}{(n-y)-k} \\
&= (-1)^{n-x-y} \sum_{k=0}^{n-y} \dbinom{k}{x}\dbinom{n-k}{(n-y)-k}.
\end{aligned} \tag{255}$$

But every $k \in \{0,1,\ldots,n-y\}$ satisfies

$$\dbinom{n-k}{(n-y)-k} = \dbinom{n-k}{y} \tag{256}$$

$x+y$). Hence, $\dbinom{n-k}{y} = 0$ (by (231), applied to $n-k$ and y instead of m and n). Therefore, $\dbinom{k}{x}\underbrace{\dbinom{n-k}{y}}_{=0} = 0$, qed.

[119]. Thus, (255) yields

$$\binom{(-y-1)-x-1}{(n-y)-x} = (-1)^{n-x-y} \sum_{k=0}^{n-y} \binom{k}{x} \underbrace{\binom{n-k}{(n-y)-k}}_{\substack{=\binom{n-k}{y} \\ \text{(by (256))}}} = (-1)^{n-x-y} \sum_{k=0}^{n-y} \binom{k}{x} \binom{n-k}{y}.$$

Compared with

$$\binom{(-y-1)-x-1}{(n-y)-x} = \underbrace{(-1)^{(n-y)-x}}_{\substack{=(-1)^{n-x-y} \\ \text{(since } (n-y)-x=n-x-y)}} \underbrace{\binom{((n-y)-x)-((-y-1)-x-1)-1}{(n-y)-x}}_{\substack{=\binom{n+1}{n-x-y} \\ \text{(since } ((n-y)-x)-((-y-1)-x-1)-1=n+1 \\ \text{and } (n-y)-x=n-x-y)}}$$

$$\left(\begin{array}{c}\text{by (237), applied to } (-y-1)-x-1 \text{ and } (n-y)-x \\ \text{instead of } m \text{ and } n\end{array}\right)$$

$$= (-1)^{n-x-y} \binom{n+1}{n-x-y},$$

this yields

$$(-1)^{n-x-y} \binom{n+1}{n-x-y} = (-1)^{n-x-y} \sum_{k=0}^{n-y} \binom{k}{x} \binom{n-k}{y}.$$

We can cancel $(-1)^{n-x-y}$ from this equality (because $(-1)^{n-x-y} \neq 0$). As a result, we obtain

$$\binom{n+1}{n-x-y} = \sum_{k=0}^{n-y} \binom{k}{x} \binom{n-k}{y}. \tag{257}$$

But $0 \leq n-y$ (since $n-y \in \mathbb{N}$) and $n-y \leq n$ (since $y \in \mathbb{N}$). Also, every $k \in \{n-y+1, n-y+2, \ldots, n\}$ satisfies

$$\binom{n-k}{y} = 0 \tag{258}$$

[119] *Proof of (256):* Let $k \in \{0, 1, \ldots, n-y\}$. Then, $k \in \mathbb{N}$ and $n-y \geq k$. From $n-y \geq k$, we obtain $n \geq y+k$, so that $n-k \geq y$. Thus, $n-k \geq y \geq 0$, so that $n-k \in \mathbb{N}$. Hence, (232) (applied to $n-k$ and y instead of m and n) yields $\binom{n-k}{y} = \binom{n-k}{(n-k)-y} = \binom{n-k}{(n-y)-k}$ (since $(n-k)-y = (n-y)-k$). This proves (256).

[120]. Hence,

$$\sum_{k=0}^{n}\binom{k}{x}\binom{n-k}{y} = \sum_{k=0}^{n-y}\binom{k}{x}\binom{n-k}{y} + \sum_{k=n-y+1}^{n}\binom{k}{x}\underbrace{\binom{n-k}{y}}_{\substack{=0\\ \text{(by (258))}}}$$

$$\left(\text{since } 0 \leq n-y \leq n\right)$$

$$= \sum_{k=0}^{n-y}\binom{k}{x}\binom{n-k}{y} + \underbrace{\sum_{k=n-y+1}^{n}\binom{k}{x}0}_{=0} = \sum_{k=0}^{n-y}\binom{k}{x}\binom{n-k}{y}$$

$$= \binom{n+1}{n-x-y} \qquad \text{(by (257))}. \tag{259}$$

Finally, $n+1 \in \mathbb{N}$ and $x+y+1 \in \mathbb{N}$ (since $x \in \mathbb{N}$ and $y \in \mathbb{N}$) and $\underbrace{n}_{\geq x+y} +1 \geq x+y+1$. Hence, (232) (applied to $n+1$ and $x+y+1$ instead of m and n) yields

$$\binom{n+1}{x+y+1} = \binom{n+1}{(n+1)-(x+y+1)} = \binom{n+1}{n-x-y}$$

(since $(n+1)-(x+y+1) = n-x-y$). Comparing this with (259), we obtain

$$\binom{n+1}{x+y+1} = \sum_{k=0}^{n}\binom{k}{x}\binom{n-k}{y}.$$

Thus, Proposition 3.32 **(f)** is proven in Case 2.

We have now proven Proposition 3.32 **(f)** in both Cases 1 and 2; thus, Proposition 3.32 **(f)** always holds.

(g) Let $x \in \mathbb{Z}$ and $y \in \mathbb{N}$ and $n \in \mathbb{N}$ be such that $x+y \geq 0$ and $n \geq x$. We have $x+y \in \mathbb{N}$ (since $x+y \geq 0$). We must be in one of the following two cases:

Case 1: We have $x+y < n$.

Case 2: We have $x+y \geq n$.

Let us first consider Case 1. In this case, we have $x+y < n$. Thus, $\binom{x+y}{n} = 0$ (by (231), applied to $m = x+y$). But every $k \in \{0,1,\ldots,x+y\}$ satisfies $\binom{y}{n+k-x} =$

[120]*Proof of (258):* Let $k \in \{n-y+1, n-y+2, \ldots, n\}$. Then, $k \leq n$ and $k > n-y$. Hence, $n-k \in \mathbb{N}$ (since $k \leq n$) and $n - \underbrace{k}_{>n-y} < n-(n-y) = y$. Therefore, (231) (applied to $n-k$ and y instead of m and n) yields $\binom{n-k}{y} = 0$. This proves (258).

0 [121]. Thus, $\sum_{k=0}^{x+y} \binom{x}{k} \underbrace{\binom{y}{n+k-x}}_{=0} = \sum_{k=0}^{x+y} \binom{x}{k} 0 = 0$. Compared with $\binom{x+y}{n} = 0$, this yields $\binom{x+y}{n} = \sum_{k=0}^{x+y} \binom{x}{k} \binom{y}{n+k-x}$. Thus, Proposition 3.32 **(g)** is proven in Case 1.

Let us now consider Case 2. In this case, we have $x+y \geq n$. Hence, $\binom{x+y}{n} = \binom{x+y}{x+y-n}$ (by (232), applied to $m = x+y$). Also, $x+y-n \in \mathbb{N}$ (since $x+y \geq n$). Therefore, Proposition 3.32 **(a)** (applied to $x+y-n$ instead of n) yields

$$\binom{x+y}{x+y-n} = \sum_{k=0}^{x+y-n} \binom{x}{k} \binom{y}{x+y-n-k}.$$

Since $\binom{x+y}{n} = \binom{x+y}{x+y-n}$, this rewrites as

$$\binom{x+y}{n} = \sum_{k=0}^{x+y-n} \binom{x}{k} \binom{y}{x+y-n-k}. \tag{260}$$

But every $k \in \{0, 1, \ldots, x+y-n\}$ satisfies $\binom{y}{x+y-n-k} = \binom{y}{n+k-x}$ [122]. Hence, (260) becomes

$$\binom{x+y}{n} = \sum_{k=0}^{x+y-n} \binom{x}{k} \underbrace{\binom{y}{x+y-n-k}}_{=\binom{y}{n+k-x}} = \sum_{k=0}^{x+y-n} \binom{x}{k} \binom{y}{n+k-x}. \tag{261}$$

On the other hand, we have $0 \leq n \leq x+y$ and thus $0 \leq x+y-n \leq x+y$. But every $k \in \mathbb{N}$ satisfying $k > x+y-n$ satisfies

$$\binom{y}{n+k-x} = 0 \tag{262}$$

[121] *Proof.* Let $k \in \{0, 1, \ldots, x+y\}$. Then, $k \geq 0$, so that $n + \underbrace{k}_{\geq 0} - x \geq n - x > y$ (since $n > x+y$ (since $x+y < n$)). In other words, $y < n+k-x$. Also, $n+k-x > y \geq 0$, so that $n+k-x \in \mathbb{N}$. Hence, $\binom{y}{n+k-x} = 0$ (by (231), applied to y and $n+k-x$ instead of m and n). Qed.

[122] *Proof.* Let $k \in \{0, 1, \ldots, x+y-n\}$. Then, $0 \leq k \leq x+y-n$. Hence, $x+y-n \geq k$, so that $x+y-n-k \in \mathbb{N}$. Also, $y \geq x+y-n-k$ (since $y-(x+y-n-k) = \underbrace{n}_{\geq x} + \underbrace{k}_{\geq 0} - x \geq x+0-x = 0$). Therefore, (232) (applied to y and $x+y-n-k$ instead of m and n) yields $\binom{y}{x+y-n-k} = \binom{y}{y-(x+y-n-k)} = \binom{y}{n+k-x}$ (since $y-(x+y-n-k) = n+k-x$), qed.

[123]. Hence,

$$\sum_{k=0}^{x+y} \binom{x}{k} \binom{y}{n+k-x}$$
$$= \sum_{k=0}^{x+y-n} \binom{x}{k} \binom{y}{n+k-x} + \sum_{k=(x+y-n)+1}^{x+y} \binom{x}{k} \underbrace{\binom{y}{n+k-x}}_{\substack{=0 \\ \text{(by (262) (since } k \geq (x+y-n)+1 > x+y-n))}}$$
$$\left(\text{since } 0 \leq x+y-n \leq x+y\right)$$
$$= \sum_{k=0}^{x+y-n} \binom{x}{k} \binom{y}{n+k-x} + \underbrace{\sum_{k=(x+y-n)+1}^{x+y} \binom{x}{k} 0}_{=0} = \sum_{k=0}^{x+y-n} \binom{x}{k} \binom{y}{n+k-x}.$$

Compared with (261), this yields

$$\binom{x+y}{n} = \sum_{k=0}^{x+y} \binom{x}{k} \binom{y}{n+k-x}.$$

This proves Proposition 3.32 **(g)** in Case 2.

Proposition 3.32 **(g)** is thus proven in each of the two Cases 1 and 2. Therefore, Proposition 3.32 **(g)** holds in full generality. □

Remark 3.34. The proof of Proposition 3.32 given above illustrates a useful technique: the use of upper negation (i.e., the equality (237)) to transform one equality into another. In a nutshell,

- we have proven Proposition 3.32 **(d)** by applying Proposition 3.32 **(a)** to $-y$ instead of y, and then rewriting the result using upper negation;
- we have proven Proposition 3.32 **(e)** by applying Proposition 3.32 **(a)** to y, $-x-1$ and $n-x$ instead of x, y and n, and then rewriting the resulting identity using upper negation;
- we have proven Proposition 3.32 **(f)** by applying Proposition 3.32 **(e)** to $-y-1$ and $n-y$ instead of y and n, and rewriting the resulting identity using upper negation.

Thus, by substitution and rewriting using upper negation, one single equality (namely, Proposition 3.32 **(a)**) has morphed into three other equalities. Note, in particular, that no negative numbers appear in Proposition 3.32 **(f)**, but yet we proved it by substituting negative values for x and y.

[123] *Proof.* Let $k \in \mathbb{N}$ be such that $k > x+y-n$. Then, $n + \underbrace{k}_{> x+y-n} - x > n + (x+y-n) - x = y$. In other words, $y < n+k-x$. Also, $n+k-x > y \geq 0$, so that $n+k-x \in \mathbb{N}$. Hence, (231) (applied to y and $n+k-x$ instead of m and n) yields $\binom{y}{n+k-x} = 0$, qed.

3.4. Further results

Exercise 3.8. Let n be a nonnegative integer. Prove that there exist **nonnegative** integers $c_{i,j}$ for all $0 \leq i \leq n$ and $0 \leq j \leq n$ such that

$$\binom{XY}{n} = \sum_{i=0}^{n} \sum_{j=0}^{n} c_{i,j} \binom{X}{i} \binom{Y}{j} \qquad (263)$$

(an equality between polynomials in two variables X and Y).

Notice that the integers $c_{i,j}$ in Exercise 3.8 can depend on the n (besides depending on i and j). We just have not included the n in the notation because it is fixed.

We shall now state two results that are used by Lee and Schiffler in their celebrated proof of positivity for cluster algebras [LeeSch13] (one of the recent breakthroughs in cluster algebra theory). Specifically, our Exercise 3.9 is (essentially) [LeeSch13, Lemma 5.11], and our Proposition 3.35 is (essentially) [LeeSch13, Lemma 5.12][124].

Exercise 3.9. Let a, b and c be three nonnegative integers. Prove that the polynomial $\binom{aX+b}{c}$ in the variable X (this is a polynomial in X of degree $\leq c$) can be written as a sum $\sum_{i=0}^{c} d_i \binom{X}{i}$ with **nonnegative** d_i.

Proposition 3.35. Let a and b be two nonnegative integers. There exist **nonnegative** integers $e_0, e_1, \ldots, e_{a+b}$ such that

$$\binom{X}{a} \binom{X}{b} = \sum_{i=0}^{a+b} e_i \binom{X}{i}$$

(an equality between polynomials in X).

First proof of Proposition 3.35. For every $N \in \mathbb{N}$, we let $[N]$ denote the N-element set $\{1, 2, \ldots, N\}$.

For every set S, we let an *S-junction* mean a pair (A, B), where A is an a-element subset of S and where B is a b-element subset of S such that $A \cup B = S$. (We do not mention a and b in our notation, because a and b are fixed.)

For example, if $a = 2$ and $b = 3$, then $(\{1,4\}, \{2,3,4\})$ is a $[4]$-junction, and $(\{2,4\}, \{1,4,6\})$ is a $\{1,2,4,6\}$-junction, but $(\{1,3\}, \{2,3,5\})$ is not a $[5]$-junction (since $\{1,3\} \cup \{2,3,5\} \neq [5]$).

[124] We say "essentially" because the X in [LeeSch13, Lemma 5.11] and in [LeeSch13, Lemma 5.12] is a variable ranging over the nonnegative integers rather than an indeterminate. But this does not make much of a difference (indeed, Lemma 3.28 **(b)** allows us to easily derive our Exercise 3.9 and Proposition 3.35 from [LeeSch13, Lemma 5.11] and [LeeSch13, Lemma 5.12], and of course the converse implication is obvious).

For every $i \in \mathbb{N}$, we let e_i be the number of all $[i]$-junctions. Then, if S is any i-element set, then

$$e_i \text{ is the number of all } S\text{-junctions} \qquad (264)$$

[125].

Now, let us show that

$$\binom{x}{a}\binom{x}{b} = \sum_{i=0}^{a+b} e_i \binom{x}{i} \qquad (265)$$

for every $x \in \mathbb{N}$.

[*Proof of (265):* Let $x \in \mathbb{N}$. How many ways are there to choose a pair (A, B) consisting of an a-element subset A of $[x]$ and a b-element subset B of $[x]$?

Let us give two different answers to this question. The first answer is the straightforward one: To choose a pair (A, B) consisting of an a-element subset A of $[x]$ and a b-element subset B of $[x]$, we need to choose an a-element subset A of $[x]$ and a b-element subset B of $[x]$. There are $\binom{x}{a}\binom{x}{b}$ total ways to do this (since there are $\binom{x}{a}$ choices for A [126], and $\binom{x}{b}$ choices for B [127], and these choices are independent). In other words, the number of all pairs (A, B) consisting of an a-element subset A of $[x]$ and a b-element subset B of $[x]$ equals $\binom{x}{a}\binom{x}{b}$.

On the other hand, here is a more imaginative procedure to choose a pair (A, B) consisting of an a-element subset A of $[x]$ and a b-element subset B of $[x]$:

1. We choose how many elements the union $A \cup B$ will have. In other words, we choose an $i \in \mathbb{N}$ that will satisfy $|A \cup B| = i$. This i must be an integer between 0 and $a + b$ (inclusive)[128].

---

[125] *Proof of (264):* Let S be any i-element set. We know that e_i is the number of all $[i]$-junctions. We want to prove that e_i is the number of all S-junctions. Roughly speaking, this is obvious, because we can "relabel the elements of S as $1, 2, \ldots, i$" (since S is an i-element set), and then the S-junctions become precisely the $[i]$-junctions.

Here is a formal way to make this argument: The sets $[i]$ and S have the same number of elements (indeed, both are i-element sets). Hence, there exists a bijection $\phi : S \to [i]$. Fix such a ϕ. Now, the S-junctions are in a 1-to-1 correspondence with the $[i]$-junctions (namely, to every S-junction (A, B) corresponds the $[i]$-junction $(\phi(A), \phi(B))$, and conversely, to every $[i]$-junction (A', B') corresponds the S-junction $(\phi^{-1}(A'), \phi^{-1}(B'))$). Hence, the number of all S-junctions equals the number of $[i]$-junctions. Since the latter number is e_i, this shows that the former number is also e_i. This proves (264).

[126] This follows from (236).

[127] Again, this follows from (236).

[128] *Proof.* Clearly, i cannot be smaller than 0. But i also cannot be larger than $a + b$ (since i will have to satisfy $i = |A \cup B| \leq \underbrace{|A|}_{=a} + \underbrace{|B|}_{=b} = a + b$). Thus, i must be an integer between 0 and $a + b$ (inclusive).

2. We choose a subset S of $[x]$, which will serve as the union $A \cup B$. This subset S must be an i-element subset of $[x]$ (because we will have $\left|\underbrace{S}_{=A\cup B}\right| = |A \cup B| = i$). Thus, there are $\binom{x}{i}$ ways to choose it (since we need to choose an i-element subset of $[x]$).

3. Now, it remains to choose the pair (A, B) itself. This pair must be a pair of subsets of $[x]$ satisfying $|A| = a$, $|B| = b$, $A \cup B = S$ and $|A \cup B| = i$. We can forget about the $|A \cup B| = i$ condition, since it automatically follows from $A \cup B = S$ (because $|S| = i$). So we need to choose a pair (A, B) of subsets of $[x]$ satisfying $|A| = a$, $|B| = b$ and $A \cup B = S$. In other words, we need to choose a pair (A, B) of subsets of S satisfying $|A| = a$, $|B| = b$ and $A \cup B = S$ [129]. In other words, we need to choose an S-junction (since this is how an S-junction was defined). This can be done in exactly e_i ways (according to (264)).

Thus, in total, there are $\sum_{i=0}^{a+b} \binom{x}{i} e_i$ ways to perform this procedure. Hence, the total number of all pairs (A, B) consisting of an a-element subset A of $[x]$ and a b-element subset B of $[x]$ equals $\sum_{i=0}^{a+b} \binom{x}{i} e_i$. But earlier, we have shown that this number is $\binom{x}{a}\binom{x}{b}$. Comparing these two results, we conclude that $\binom{x}{a}\binom{x}{b} = \sum_{i=0}^{a+b} \binom{x}{i} e_i = \sum_{i=0}^{a+b} e_i \binom{x}{i}$. Thus, (265) is proven.]

Now, we define two polynomials P and Q in the indeterminate X with rational coefficients by setting

$$P = \binom{X}{a}\binom{X}{b}; \qquad Q = \sum_{i=0}^{a+b} e_i \binom{X}{i}.$$

The equality (265) (which we have proven) states that $P(x) = Q(x)$ for all $x \in \mathbb{N}$. Thus, Lemma 3.28 **(b)** yields that $P = Q$. Recalling how P and Q are defined, we see that this rewrites as $\binom{X}{a}\binom{X}{b} = \sum_{i=0}^{a+b} e_i \binom{X}{i}$. This proves Proposition 3.35. □

Our second proof of Proposition 3.35 is algebraic, and is based on a suggestion of math.stackexchange user tcamps in a comment on question #1342384. It proceeds by way of the following, more explicit result:

[129] Here, we have replaced "subsets of $[x]$" by "subsets of S", because the condition $A \cup B = S$ forces A and B to be subsets of S.

Proposition 3.36. Let a and b be two nonnegative integers. Then,

$$\binom{X}{a}\binom{X}{b} = \sum_{i=a}^{a+b} \binom{i}{a}\binom{a}{a+b-i}\binom{X}{i}.$$

Let us also state the analogue of this proposition in which the indeterminate X is replaced by a rational number m:

Proposition 3.37. Let a and b be two nonnegative integers. Let $m \in \mathbb{Q}$. Then,

$$\binom{m}{a}\binom{m}{b} = \sum_{i=a}^{a+b} \binom{i}{a}\binom{a}{a+b-i}\binom{m}{i}.$$

Proof of Proposition 3.37. Theorem 3.29 (applied to b, $m-a$ and a instead of n, x and y) yields

$$\binom{(m-a)+a}{b} = \sum_{k=0}^{b} \binom{m-a}{k}\binom{a}{b-k}.$$

Since $(m-a)+a = m$, this rewrites as

$$\binom{m}{b} = \sum_{k=0}^{b} \binom{m-a}{k}\binom{a}{b-k} = \sum_{i=a}^{a+b} \binom{m-a}{i-a} \underbrace{\binom{a}{b-(i-a)}}_{\substack{=\binom{a}{a+b-i} \\ \text{(since } b-(i-a)=a+b-i\text{)}}}$$

(here, we substituted $i-a$ for k in the sum)

$$= \sum_{i=a}^{a+b} \binom{m-a}{i-a}\binom{a}{a+b-i}.$$

Multiplying both sides of this identity with $\binom{m}{a}$, we obtain

$$\binom{m}{a}\binom{m}{b} = \binom{m}{a}\sum_{i=a}^{a+b} \binom{m-a}{i-a}\binom{a}{a+b-i} = \sum_{i=a}^{a+b} \underbrace{\binom{m}{a}\binom{m-a}{i-a}}_{\substack{=\binom{m}{i}\binom{i}{a} \\ \text{(by Proposition 3.23)}}} \binom{a}{a+b-i}$$

$$= \sum_{i=a}^{a+b} \binom{m}{i}\binom{i}{a}\binom{a}{a+b-i} = \sum_{i=a}^{a+b} \binom{i}{a}\binom{a}{a+b-i}\binom{m}{i}.$$

This proves Proposition 3.37. □

Proof of Proposition 3.36. To obtain a proof of Proposition 3.36, replace every appearance of "m" by "X" in the above proof of Proposition 3.37. (Of course, this requires knowing that Theorem 3.29 holds when x and y are polynomials rather than numbers. But this is true, because both proofs that we gave for Theorem 3.29 still apply in this case.) □

Second proof of Proposition 3.35. Let us define $a+b+1$ nonnegative integers $e_0, e_1, \ldots, e_{a+b}$ by

$$e_i = \begin{cases} \dbinom{i}{a}\dbinom{a}{a+b-i}, & \text{if } i \geq a; \\ 0, & \text{otherwise} \end{cases} \qquad \text{for all } i \in \{0, 1, \ldots, a+b\}. \tag{266}$$

Then,

$$\begin{aligned} \sum_{i=0}^{a+b} e_i \binom{X}{i} &= \sum_{i=a}^{a+b} \binom{i}{a}\binom{a}{a+b-i}\binom{X}{i} && \left(\text{by our definition of } e_0, e_1, \ldots, e_{a+b}\right) \\ &= \binom{X}{a}\binom{X}{b} && (\text{by Proposition 3.36}). \end{aligned}$$

Thus, Proposition 3.35 is proven again. □

Remark 3.38. Comparing our two proofs of Proposition 3.35, it is natural to suspect that the $e_0, e_1, \ldots, e_{a+b}$ defined in the First proof are identical with the $e_0, e_1, \ldots, e_{a+b}$ defined in the Second proof. This actually follows from general principles (namely, from the word "unique" in Proposition 3.27 **(a)**), but there is also a simple combinatorial reason. Namely, let $i \in \{0, 1, \ldots, a+b\}$. We shall show that the e_i defined in the First proof equals the e_i defined in the Second proof.

The e_i defined in the First proof is the number of all $[i]$-junctions. An $[i]$-junction is a pair (A, B), where A is an a-element subset of $[i]$ and where B is a b-element subset of $[i]$ such that $A \cup B = [i]$. Here is a way to construct an $[i]$-junction:

- First, we pick the set A. There are $\dbinom{i}{a}$ ways to do this, since A has to be an a-element subset of the i-element set $[i]$.
- Then, we pick the set B. This has to be a b-element subset of the i-element set $[i]$ satisfying $A \cup B = [i]$. The equality $A \cup B = [i]$ means that B has to contain the $i-a$ elements of $[i] \setminus A$; but the remaining $b-(i-a) = a+b-i$ elements of B can be chosen arbitrarily among the a elements of A. Thus, there are $\dbinom{a}{a+b-i}$ ways to choose B (since we have to choose $a+b-i$ elements of B among the a elements of A).

Thus, the number of all $[i]$-junctions is $\binom{i}{a}\binom{a}{a+b-i}$. This can be rewritten in the form $\begin{cases} \binom{i}{a}\binom{a}{a+b-i}, & \text{if } i \geq a; \\ 0, & \text{otherwise} \end{cases}$ (because if $i < a$, then $\binom{i}{a} = 0$ and thus $\binom{i}{a}\binom{a}{a+b-i} = 0$). Thus, we have shown that the number of all $[i]$-junctions is $\begin{cases} \binom{i}{a}\binom{a}{a+b-i}, & \text{if } i \geq a; \\ 0, & \text{otherwise} \end{cases}$. In other words, the e_i defined in the First proof equals the e_i defined in the Second proof.

Here is an assortment of other identities that involve binomial coefficients:

Proposition 3.39. **(a)** Every $x \in \mathbb{Z}$, $y \in \mathbb{Z}$ and $n \in \mathbb{N}$ satisfy $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

(b) Every $n \in \mathbb{N}$ satisfies $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

(c) Every $n \in \mathbb{N}$ satisfies $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = \begin{cases} 1, & \text{if } n = 0; \\ 0, & \text{if } n \neq 0 \end{cases}$.

(d) Every $n \in \mathbb{Z}$, $i \in \mathbb{N}$ and $a \in \mathbb{N}$ satisfying $i \geq a$ satisfy $\binom{n}{i}\binom{i}{a} = \binom{n}{a}\binom{n-a}{i-a}$.

(e) Every $n \in \mathbb{N}$ and $m \in \mathbb{Z}$ satisfy $\sum_{i=0}^{n} \binom{n}{i}\binom{m+i}{n} = \sum_{i=0}^{n} \binom{n}{i}\binom{m}{i} 2^i$.

(f) Every $a \in \mathbb{N}$, $b \in \mathbb{N}$ and $x \in \mathbb{Z}$ satisfy $\sum_{i=0}^{b} \binom{a}{i}\binom{b}{i}\binom{x+i}{a+b} = \binom{x}{a}\binom{x}{b}$.

(g) Every $a \in \mathbb{N}$, $b \in \mathbb{N}$ and $x \in \mathbb{Z}$ satisfy $\sum_{i=0}^{b} \binom{a}{i}\binom{b}{i}\binom{a+b+x-i}{a+b} = \binom{a+x}{a}\binom{b+x}{b}$.

(I have learnt parts **(e)** and **(f)** of Proposition 3.39 from AoPS, but they are fairly classical results. Part **(e)** is equivalent to a claim in [Comtet74, Chapter I, Exercise 21]. Part **(f)** is [Riorda68, §1.4, (10)]. Part **(g)** is a restatement of [Gould10, (6.93)].)

Proof of Proposition 3.39. **(a)** Proposition 3.39 **(a)** is clearly a particular case of Proposition 3.21.

(b) Let $n \in \mathbb{N}$. Applying Proposition 3.39 **(a)** to $x = 1$ and $y = 1$, we obtain

$$(1+1)^n = \sum_{k=0}^{n} \binom{n}{k} \underbrace{1^k}_{=1} \underbrace{1^{n-k}}_{=1} = \sum_{k=0}^{n} \binom{n}{k},$$

thus

$$\sum_{k=0}^{n} \binom{n}{k} = \Bigg(\underbrace{1+1}_{=2}\Bigg)^n = 2^n.$$

This proves Proposition 3.39 **(b)**.

(c) Let $n \in \mathbb{N}$. Applying Proposition 3.39 **(a)** to $x = -1$ and $y = 1$, we obtain

$$(-1+1)^n = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \underbrace{1^{n-k}}_{=1} = \sum_{k=0}^{n} \binom{n}{k} (-1)^k = \sum_{k=0}^{n} (-1)^k \binom{n}{k},$$

thus

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = \Bigg(\underbrace{-1+1}_{=0}\Bigg)^n = 0^n = \begin{cases} 1, & \text{if } n = 0; \\ 0, & \text{if } n \neq 0 \end{cases}.$$

This proves Proposition 3.39 **(c)**.

(d) Let $n \in \mathbb{Z}$, $i \in \mathbb{N}$ and $a \in \mathbb{N}$ be such that $i \geq a$. Proposition 3.23 (applied to n instead of m) yields $\binom{n}{i}\binom{i}{a} = \binom{n}{a}\binom{n-a}{i-a}$. This proves Proposition 3.39 **(d)**.

(e) Let $n \in \mathbb{N}$ and $m \in \mathbb{Z}$. Clearly, every $p \in \mathbb{N}$ satisfies

$$\sum_{i=0}^{p} \binom{p}{i} = \sum_{k=0}^{p} \binom{p}{k} \qquad \text{(here, we renamed the summation index } i \text{ as } k\text{)}$$
$$= 2^p \tag{267}$$

(by Proposition 3.39 **(b)**, applied to p instead of n).

Now, let $i \in \{0, 1, \ldots, n\}$. Applying Proposition 3.32 **(a)** to $x = i$ and $y = m$, we obtain

$$\binom{i+m}{n} = \sum_{k=0}^{n} \binom{i}{k} \binom{m}{n-k}$$
$$= \sum_{k=0}^{i} \binom{i}{k} \binom{m}{n-k} + \sum_{k=i+1}^{n} \underbrace{\binom{i}{k}}_{\substack{=0 \\ \text{(by (231), applied to } i \text{ and } k \\ \text{instead of } m \text{ and } n \text{ (since } i<k \\ \text{(because } k \geq i+1 > i\text{)))}}} \binom{m}{n-k} \qquad (\text{since } 0 \leq i \leq n)$$
$$= \sum_{k=0}^{i} \binom{i}{k} \binom{m}{n-k} + \underbrace{\sum_{k=i+1}^{n} 0 \binom{m}{n-k}}_{=0} = \sum_{k=0}^{i} \binom{i}{k} \binom{m}{n-k}. \tag{268}$$

Now, let us forget that we fixed i. We thus have proven (268) for every $i \in \{0, 1, \ldots, n\}$. Now,

$$\begin{aligned}
\sum_{i=0}^{n} \binom{n}{i} \underbrace{\binom{m+i}{n}}_{\substack{=\binom{i+m}{n}=\sum\limits_{k=0}^{i}\binom{i}{k}\binom{m}{n-k} \\ \text{(by (268))}}}
&= \sum_{i=0}^{n} \binom{n}{i} \left(\sum_{k=0}^{i} \binom{i}{k} \binom{m}{n-k} \right) = \underbrace{\sum_{i=0}^{n} \sum_{k=0}^{i}}_{=\sum\limits_{k=0}^{n}\sum\limits_{i=k}^{n}} \binom{n}{i} \binom{i}{k} \binom{m}{n-k} \\
&= \sum_{k=0}^{n} \sum_{i=k}^{n} \underbrace{\binom{n}{i}\binom{i}{k}}_{\substack{=\binom{n}{k}\binom{n-k}{i-k} \\ \text{(by Proposition 3.39 (d),} \\ \text{applied to } a=k \text{ (since } i \geq k))}} \binom{m}{n-k} \\
&= \sum_{k=0}^{n} \sum_{i=k}^{n} \binom{n}{k} \binom{n-k}{i-k} \binom{m}{n-k} = \sum_{k=0}^{n} \binom{n}{k} \binom{m}{n-k} \sum_{i=k}^{n} \binom{n-k}{i-k} \\
&= \sum_{k=0}^{n} \underbrace{\binom{n}{k}}_{\substack{=\binom{n}{n-k} \\ \text{(by (232), applied to } n \text{ and } k \\ \text{instead of } m \text{ and } n \text{ (since } n \geq k))}} \binom{m}{n-k} \sum_{i=0}^{n-k} \binom{n-k}{i} \\
&\qquad \left(\text{here, we have substituted } i \text{ for } i-k \text{ in the second sum}\right) \\
&= \sum_{k=0}^{n} \binom{n}{n-k} \binom{m}{n-k} \sum_{i=0}^{n-k} \binom{n-k}{i} = \sum_{k=0}^{n} \binom{n}{k} \binom{m}{k} \underbrace{\sum_{i=0}^{k} \binom{k}{i}}_{\substack{=2^k \\ \text{(by (267), applied to } p=k)}} \\
&\qquad \left(\text{here, we have substituted } k \text{ for } n-k \text{ in the first sum}\right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \binom{m}{k} 2^k = \sum_{i=0}^{n} \binom{n}{i} \binom{m}{i} 2^i
\end{aligned}$$

(here, we have renamed the summation index k as i). This proves Proposition 3.39 **(e)**.

(f) Let $a \in \mathbb{N}$, $b \in \mathbb{N}$ and $x \in \mathbb{Z}$. Proposition 3.37 (applied to $m = x$) yields

$$\binom{x}{a} \binom{x}{b} = \sum_{i=a}^{a+b} \binom{i}{a} \binom{a}{a+b-i} \binom{x}{i}. \tag{269}$$

Clearly,

$$\begin{aligned}
&\sum_{i=0}^{a+b} \binom{a}{i} \binom{b}{i} \binom{x+i}{a+b} \\
&= \sum_{i=0}^{b} \binom{a}{i} \binom{b}{i} \binom{x+i}{a+b} + \sum_{i=b+1}^{a+b} \binom{a}{i} \underbrace{\binom{b}{i}}_{\substack{=0 \\ \text{(by (231), applied to } b \text{ and } i \\ \text{instead of } m \text{ and } n \text{ (since } b<i \\ \text{(because } i \geq b+1 > b)))}} \binom{x+i}{a+b} \\
&\qquad (\text{since } 0 \leq b \leq a+b) \\
&= \sum_{i=0}^{b} \binom{a}{i} \binom{b}{i} \binom{x+i}{a+b} + \underbrace{\sum_{i=b+1}^{a+b} \binom{a}{i} 0 \binom{x+i}{a+b}}_{=0} \\
&= \sum_{i=0}^{b} \binom{a}{i} \binom{b}{i} \binom{x+i}{a+b}.
\end{aligned} \tag{270}$$

For every $i \in \{0, 1, \ldots, b\}$, we have

$$\binom{x+i}{a+b} = \sum_{k=0}^{a+b} \binom{x}{k} \binom{i}{a+b-k}. \tag{271}$$

(This follows from Theorem 3.29 (applied to $a+b$ and i instead of n and y).) Hence,

$$\begin{aligned}
&\sum_{i=0}^{a+b} \binom{a}{i} \binom{b}{i} \underbrace{\binom{x+i}{a+b}}_{\substack{= \sum_{k=0}^{a+b} \binom{x}{k} \binom{i}{a+b-k} \\ \text{(by (271))}}} \\
&= \sum_{i=0}^{a+b} \binom{a}{i} \binom{b}{i} \sum_{k=0}^{a+b} \binom{x}{k} \binom{i}{a+b-k} \\
&= \underbrace{\sum_{i=0}^{a+b} \sum_{k=0}^{a+b}}_{= \sum_{k=0}^{a+b} \sum_{i=0}^{a+b}} \binom{a}{i} \underbrace{\binom{b}{i} \binom{x}{k} \binom{i}{a+b-k}}_{= \binom{i}{a+b-k} \binom{b}{i} \binom{x}{k}} \\
&= \sum_{k=0}^{a+b} \sum_{i=0}^{a+b} \binom{a}{i} \binom{i}{a+b-k} \binom{b}{i} \binom{x}{k} = \sum_{k=0}^{a+b} \binom{x}{k} \sum_{i=0}^{a+b} \binom{a}{i} \binom{i}{a+b-k} \binom{b}{i}.
\end{aligned}$$

Compared with (270), this yields

$$\sum_{i=0}^{b}\binom{a}{i}\binom{b}{i}\binom{x+i}{a+b} = \sum_{k=0}^{a+b}\binom{x}{k}\sum_{i=0}^{a+b}\binom{a}{i}\binom{i}{a+b-k}\binom{b}{i}. \tag{272}$$

But for every $k \in \{0,1,\ldots,a+b\}$, we have

$$\sum_{i=0}^{a+b}\binom{a}{i}\binom{i}{a+b-k}\binom{b}{i} = \binom{a}{a+b-k}\sum_{j=0}^{k}\binom{k-b}{k-j}\binom{b}{a+b-j}. \tag{273}$$

[*Proof of (273):* Let $k \in \{0,1,\ldots,a+b\}$. Then, $a+b-k \in \{0,1,\ldots,a+b\}$, so that

$0 \leq a+b-k \leq a+b$. Now,

$$\begin{aligned}
&\sum_{i=0}^{a+b} \binom{a}{i} \binom{i}{a+b-k} \binom{b}{i} \\
&= \sum_{i=0}^{(a+b-k)-1} \binom{a}{i} \underbrace{\binom{i}{a+b-k}}_{\substack{=0 \\ \text{(by (231), applied to } i \text{ and} \\ a+b-k \text{ instead of } m \text{ and } n \\ \text{(since } i \leq (a+b-k)-1 < a+b-k))}} \binom{b}{i} + \sum_{i=a+b-k}^{a+b} \binom{a}{i} \binom{i}{a+b-k} \binom{b}{i} \\
&\qquad \left(\text{since } 0 \leq a+b-k \leq a+b\right) \\
&= \underbrace{\sum_{i=0}^{(a+b-k)-1} \binom{a}{i} 0 \binom{b}{i}}_{=0} + \sum_{i=a+b-k}^{a+b} \binom{a}{i} \binom{i}{a+b-k} \binom{b}{i} \\
&= \sum_{i=a+b-k}^{a+b} \underbrace{\binom{a}{i} \binom{i}{a+b-k}}_{\substack{= \binom{a}{a+b-k} \binom{a-(a+b-k)}{i-(a+b-k)} \\ \text{(by Proposition 3.39 } \textbf{(d)} \text{, applied to} \\ a \text{ and } a+b-k \text{ instead of } n \text{ and } a \text{ (since } i \geq a+b-k))}} \binom{b}{i} \\
&= \sum_{i=a+b-k}^{a+b} \binom{a}{a+b-k} \binom{a-(a+b-k)}{i-(a+b-k)} \binom{b}{i} \\
&= \sum_{j=0}^{k} \binom{a}{a+b-k} \underbrace{\binom{a-(a+b-k)}{(a+b-j)-(a+b-k)}}_{\substack{= \binom{k-b}{k-j} \\ \text{(since } a-(a+b-k)=k-b \text{ and} \\ (a+b-j)-(a+b-k)=k-j)}} \binom{b}{a+b-j} \\
&\qquad \left(\text{here, we have substituted } a+b-j \text{ for } i \text{ in the sum}\right) \\
&= \sum_{j=0}^{k} \binom{a}{a+b-k} \binom{k-b}{k-j} \binom{b}{a+b-j} = \binom{a}{a+b-k} \sum_{j=0}^{k} \binom{k-b}{k-j} \binom{b}{a+b-j},
\end{aligned}$$

and this proves (273).]

Now, (272) becomes

$$\sum_{i=0}^{b} \binom{a}{i} \binom{b}{i} \binom{x+i}{a+b}$$
$$= \sum_{k=0}^{a+b} \binom{x}{k} \underbrace{\sum_{i=0}^{a+b} \binom{a}{i} \binom{i}{a+b-k} \binom{b}{i}}_{\substack{= \binom{a}{a+b-k} \sum_{j=0}^{k} \binom{k-b}{k-j} \binom{b}{a+b-j} \\ \text{(by (273))}}}$$
$$= \sum_{k=0}^{a+b} \binom{x}{k} \binom{a}{a+b-k} \sum_{j=0}^{k} \binom{k-b}{k-j} \binom{b}{a+b-j}$$
$$= \sum_{i=0}^{a+b} \binom{x}{i} \binom{a}{a+b-i} \sum_{j=0}^{i} \binom{i-b}{i-j} \binom{b}{a+b-j} \qquad (274)$$

(here, we renamed the summation index k as i in the first sum).

Furthermore, every $i \in \{0, 1, \ldots, a+b\}$ satisfies

$$\sum_{j=0}^{i} \binom{i-b}{i-j} \binom{b}{a+b-j} = \binom{i}{a}. \qquad (275)$$

[*Proof of (275):* Let $i \in \{0, 1, \ldots, a+b\}$. Thus, $0 \leq i \leq a+b$. We have

$$\sum_{j=0}^{i} \binom{i-b}{i-j} \binom{b}{a+b-j} = \sum_{k=0}^{i} \underbrace{\binom{i-b}{i-(i-k)}}_{\substack{= \binom{i-b}{k} \\ \text{(since } i-(i-k)=k)}} \underbrace{\binom{b}{a+b-(i-k)}}_{\substack{= \binom{b}{(a+b)+k-i} \\ \text{(since } a+b-(i-k)=(a+b)+k-i)}}$$

(here, we have substituted $i-k$ for j in the sum)

$$= \sum_{k=0}^{i} \binom{i-b}{k} \binom{b}{(a+b)+k-i}. \qquad (276)$$

On the other hand, we have $b \in \mathbb{N}$, $(i-b)+b = i \geq 0$ and $a \geq i-b$ (since $a+b \geq i$). Therefore, we can apply Proposition 3.32 **(g)** to $i-b$, b and a instead of

x, y and n. As a result, we obtain

$$\binom{(i-b)+b}{a} = \sum_{k=0}^{(i-b)+b} \binom{i-b}{k} \underbrace{\binom{b}{a+k-(i-b)}}_{\substack{=\binom{b}{(a+b)+k-i} \\ \text{(since } a+k-(i-b)=(a+b)+k-i)}}$$
$$= \sum_{k=0}^{(i-b)+b} \binom{i-b}{k} \binom{b}{(a+b)+k-i}.$$

Since $(i-b)+b=i$, this rewrites as

$$\binom{i}{a} = \sum_{k=0}^{i} \binom{i-b}{k} \binom{b}{(a+b)+k-i}.$$

Compared with (276), this yields

$$\sum_{j=0}^{i} \binom{i-b}{i-j} \binom{b}{a+b-j} = \binom{i}{a}.$$

This proves (275).]

Hence, (274) becomes

$$\begin{aligned}
&\sum_{i=0}^{b}\binom{a}{i}\binom{b}{i}\binom{x+i}{a+b}\\
&=\sum_{i=0}^{a+b}\binom{x}{i}\binom{a}{a+b-i}\underbrace{\sum_{j=0}^{i}\binom{i-b}{i-j}\binom{b}{a+b-j}}_{\substack{=\binom{i}{a}\\ \text{(by (275))}}}\\
&=\sum_{i=0}^{a+b}\underbrace{\binom{x}{i}\binom{a}{a+b-i}\binom{i}{a}}_{=\binom{i}{a}\binom{a}{a+b-i}\binom{x}{i}}=\sum_{i=0}^{a+b}\binom{i}{a}\binom{a}{a+b-i}\binom{x}{i}\\
&=\sum_{i=0}^{a-1}\underbrace{\binom{i}{a}}_{\substack{=0\\ \text{(by (231), applied to } i \text{ and } a\\ \text{instead of } m \text{ and } n \text{ (since } i<a))}}\binom{a}{a+b-i}\binom{x}{i}+\sum_{i=a}^{a+b}\binom{i}{a}\binom{a}{a+b-i}\binom{x}{i}\\
&\qquad(\text{since } 0\le a\le a+b)\\
&=\underbrace{\sum_{i=0}^{a-1}0\binom{a}{a+b-i}\binom{x}{i}}_{=0}+\sum_{i=a}^{a+b}\binom{i}{a}\binom{a}{a+b-i}\binom{x}{i}\\
&=\sum_{i=a}^{a+b}\binom{i}{a}\binom{a}{a+b-i}\binom{x}{i}=\binom{x}{a}\binom{x}{b}\qquad\text{(by (269))}.
\end{aligned}$$

This proves Proposition 3.39 **(f)**.

(g) Let $a\in\mathbb{N}$, $b\in\mathbb{N}$ and $x\in\mathbb{Z}$. From (237) (applied to $m=-x-1$ and $n=a$), we obtain $\binom{-x-1}{a}=(-1)^a\binom{a-(-x-1)-1}{a}=(-1)^a\binom{a+x}{a}$ (since $a-(-x-1)-1=a+x$). The same argument (applied to b instead of a) shows that $\binom{-x-1}{b}=(-1)^b\binom{b+x}{b}$.

Now, Proposition 3.39 **(f)** (applied to $-x-1$ instead of x) shows that

$$\begin{aligned}
\sum_{i=0}^{b} \binom{a}{i}\binom{b}{i}\binom{(-x-1)+i}{a+b} &= \underbrace{\binom{-x-1}{a}}_{=(-1)^a\binom{a+x}{a}} \underbrace{\binom{-x-1}{b}}_{=(-1)^b\binom{b+x}{b}} \\
&= (-1)^a \binom{a+x}{a} (-1)^b \binom{b+x}{b} \\
&= \underbrace{(-1)^a (-1)^b}_{=(-1)^{a+b}} \binom{a+x}{a}\binom{b+x}{b} \\
&= (-1)^{a+b} \binom{a+x}{a}\binom{b+x}{b}. \qquad (277)
\end{aligned}$$

But every $i \in \{0, 1, \ldots, b\}$ satisfies

$$\begin{aligned}
\binom{(-x-1)+i}{a+b} &= (-1)^{a+b} \binom{a+b-((-x-1)+i)-1}{a+b} \\
&\qquad \left(\text{by (237), applied to } m = (-x-1)+i \text{ and } n = a+b\right) \\
&= (-1)^{a+b} \binom{a+b+x-i}{a+b} \\
&\qquad \left(\text{since } a+b-((-x-1)+i)-1 = a+b+x-i\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{i=0}^{b} \binom{a}{i}\binom{b}{i} \underbrace{\binom{(-x-1)+i}{a+b}}_{=(-1)^{a+b}\binom{a+b+x-i}{a+b}} \\
&= \sum_{i=0}^{b} \binom{a}{i}\binom{b}{i} (-1)^{a+b} \binom{a+b+x-i}{a+b} = (-1)^{a+b} \sum_{i=0}^{b} \binom{a}{i}\binom{b}{i}\binom{a+b+x-i}{a+b}.
\end{aligned}$$

Comparing this with (277), we obtain

$$(-1)^{a+b} \sum_{i=0}^{b} \binom{a}{i}\binom{b}{i}\binom{a+b+x-i}{a+b} = (-1)^{a+b} \binom{a+x}{a}\binom{b+x}{b}.$$

We can cancel $(-1)^{a+b}$ from this equality (since $(-1)^{a+b} \neq 0$), and thus obtain $\sum_{i=0}^{b} \binom{a}{i}\binom{b}{i}\binom{a+b+x-i}{a+b} = \binom{a+x}{a}\binom{b+x}{b}$. This proves Proposition 3.39 **(g)**.

□

Many more examples of equalities with binomial coefficients, as well as advanced tactics for proving such equalities, can be found in [GrKnPa94, Chapter 5].

Exercise 3.10. Let $n \in \mathbb{Q}$, $a \in \mathbb{N}$ and $b \in \mathbb{N}$.

(a) Prove that every integer $j \geq a$ satisfies

$$\binom{n}{j}\binom{j}{a}\binom{n-j}{b} = \binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{j-a}.$$

(b) Compute the sum $\sum_{j=a}^{n} \binom{n}{j}\binom{j}{a}\binom{n-j}{b}$ for every integer $n \geq a$. (The result should contain no summation signs.)

3.5. The principle of inclusion and exclusion

We shall next discuss the *principle of inclusion and exclusion*, and some of its generalizations. This is a crucial result in combinatorics, which can help both in answering enumerative questions (i.e., questions of the form "how many objects of a given kind satisfy a certain set of properties") and in proving combinatorial identities (such as, to give a simple example, Proposition 3.39 **(c)**, but also various deeper results). We shall not dwell on the applications of this principle; the reader can easily find them in textbooks on enumerative combinatorics (such as [Aigner07, §5.1] or [Galvin17, §16] or [Loehr11, Chapter 4] or [Comtet74, Chapter IV] or [LeLeMe16, §15.9] or [AndFen04, Chapter 6]). We will, however, prove the principle and a few of its generalizations.

The principle itself (in one of its most basic forms) answers the following simple question: Given n finite sets $A_1, A_2, \ldots, A_n$, how do we compute the size of their union $A_1 \cup A_2 \cup \cdots \cup A_n$ if we know the sizes of all of their intersections (not just the intersection $A_1 \cap A_2 \cap \cdots \cap A_n$, but also the intersections of some of the sets only)? Let us first answer this question for specific small values of n:

- For $n = 1$, we have the tautological equality

$$|A_1| = |A_1|. \tag{278}$$

  (Only a true mathematician would begin a study with such a statement.)

- For $n = 2$, we have the known formula

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|. \tag{279}$$

  Notice that $|A_1|$ and $|A_2|$ are sizes of intersections of some of the sets A_1, A_2: Namely, A_1 is the intersection of the single set A_1, while A_2 is the intersection of the single set A_2.

- For $n = 3$, we have

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3| . \qquad (280)$$

  This is not as well-known as (279), but can be easily derived by applying (279) twice. (In fact, first apply (279) to $A_1 \cup A_2$ and A_3 instead of A_1 and A_2; then, apply (279) directly to rewrite $|A_1 \cup A_2|$.)

- For $n = 4$, we have

$$\begin{aligned} & |A_1 \cup A_2 \cup A_3 \cup A_4| \\ & = |A_1| + |A_2| + |A_3| + |A_4| \\ & \qquad - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_1 \cap A_4| - |A_2 \cap A_3| - |A_2 \cap A_4| - |A_3 \cap A_4| \\ & \qquad + |A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| + |A_2 \cap A_3 \cap A_4| \\ & \qquad - |A_1 \cap A_2 \cap A_3 \cap A_4| . \end{aligned} \qquad (281)$$

  Again, this can be derived by applying (279) many times.

The four equalities (278), (279), (280) and (281) all follow the same pattern: On the left hand side is the size $|A_1 \cup A_2 \cup \cdots \cup A_n|$ of the union $A_1 \cup A_2 \cup \cdots \cup A_n$ of all the n sets $A_1, A_2, \ldots, A_n$. On the right hand side is an "alternating sum" (i.e., a sum, but with minus signs in front of some of its addends), whose addends are the sizes of the intersections of all possible choices of **some** of the n sets $A_1, A_2, \ldots, A_n$ (except for the choice where none of the n sets are chosen; this does not have a well-defined intersection). Notice that there are $2^n - 1$ such choices, so the right hand side is an "alternating sum" of $2^n - 1$ addends. In other words, each addend on the right hand side has the form $\left|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}\right|$ for some m-tuple $(i_1, i_2, \ldots, i_m)$ of integers between 1 and n (inclusive) such that $m \geq 1$ and $i_1 < i_2 < \cdots < i_m$. The sign in front of this addend is a plus sign if m is odd, and is a minus sign if m is even. Thus, we can replace this sign by a factor of $(-1)^{m-1}$.

We can try and generalize the pattern as follows:

$$\begin{aligned} & |A_1 \cup A_2 \cup \cdots \cup A_n| \\ & = \sum_{m=1}^{n} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} (-1)^{m-1} \left|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}\right| . \end{aligned} \qquad (282)$$

Here, the summation sign "$\sum\limits_{1 \leq i_1 < i_2 < \cdots < i_m \leq n}$" on the right hand side is an abbreviation for $\sum\limits_{\substack{(i_1, i_2, \ldots, i_m) \in \{1, 2, \ldots, n\}^m; \\ i_1 < i_2 < \cdots < i_m}}$.

The equality (282) is indeed correct; it is one of several (equivalent) versions of the principle of inclusion and exclusion. For example, it appears in [Loehr11, §4.7]. We shall, however, state it differently, for the sake of better generalizability. First, we will index the intersections of **some** of the n sets $A_1, A_2, \ldots, A_n$ not by m-tuples $(i_1, i_2, \ldots, i_m) \in \{1, 2, \ldots, n\}^m$ satisfying $i_1 < i_2 < \cdots < i_m$, but rather by nonempty subsets of $\{1, 2, \ldots, n\}$. Second, our sets $A_1, A_2, \ldots, A_n$ will be labelled not by the numbers $1, 2, \ldots, n$, but rather by elements of a finite set G. This will result in a more abstract, but also more flexible version of (282).

First, we introduce some notations:

Definition 3.40. Let I be a set. For each $i \in I$, we let A_i be a set.

(a) Then, $\bigcup_{i \in I} A_i$ denotes the union of all the sets A_i for $i \in I$. This union is defined by

$$\bigcup_{i \in I} A_i = \{x \mid \text{there exists an } i \in I \text{ such that } x \in A_i\}.$$

For example, if $I = \{i_1, i_2, \ldots, i_k\}$ is a finite set, then

$$\bigcup_{i \in I} A_i = A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}.$$

Notice that $A_j \subseteq \bigcup_{i \in I} A_i$ for each $j \in I$. If I is finite, and if each of the sets A_i is finite, then their union $\bigcup_{i \in I} A_i$ is also finite.

Note that $\bigcup_{i \in \varnothing} A_i = \varnothing$.

(b) Assume that I is nonempty. Then, $\bigcap_{i \in I} A_i$ denotes the intersection of all the sets A_i for $i \in I$. This intersection is defined by

$$\bigcap_{i \in I} A_i = \{x \mid x \in A_i \text{ for all } i \in I\}.$$

For example, if $I = \{i_1, i_2, \ldots, i_k\}$ is a finite set, then

$$\bigcap_{i \in I} A_i = A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}.$$

Notice that $\bigcap_{i \in I} A_i \subseteq A_j$ for each $j \in I$. If each of the sets A_i is finite, then their intersection $\bigcap_{i \in I} A_i$ is also finite.

Caution: The intersection $\bigcap_{i \in I} A_i$ is not defined when I is empty, because this intersection would have to contain every object in the universe (which is impossible for a set).

Definition 3.41. If G is any finite set, then the sign $\sum_{I \subseteq G}$ shall mean $\sum_{I \in \mathcal{P}(G)}$, where

$\mathcal{P}(G)$ denotes the powerset of G. For example,

$$\sum_{I\subseteq\{7,8\}}\prod_{i\in I} i = \sum_{I\in\mathcal{P}(\{7,8\})}\prod_{i\in I} i = \underbrace{\prod_{i\in\varnothing} i}_{=(\text{empty product})=1} + \underbrace{\prod_{i\in\{7\}} i}_{=7} + \underbrace{\prod_{i\in\{8\}} i}_{=8} + \underbrace{\prod_{i\in\{7,8\}} i}_{=7\cdot 8}$$
$$\left(\text{since the subsets of } \{7,8\} \text{ are } \varnothing, \{7\}, \{8\}, \{7,8\}\right)$$
$$= 1+7+8+7\cdot 8.$$

We are now ready to state one of the forms of the principle:

Theorem 3.42. Let G be a finite set. For each $i \in G$, let A_i be a finite set. Then,

$$\left|\bigcup_{i\in G} A_i\right| = \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|-1}\left|\bigcap_{i\in I} A_i\right|.$$

If $G = \{1,2,\ldots,n\}$ for some $n \in \mathbb{N}$, then the formula in Theorem 3.42 is a restatement of (282) (because the nonempty subsets of $\{1,2,\ldots,n\}$ are in 1-to-1 correspondence with the m-tuples $(i_1,i_2,\ldots,i_m) \in \{1,2,\ldots,n\}^m$ satisfying $i_1 < i_2 < \cdots < i_m$ and $m \in \{1,2,\ldots,n\}$).

A statement equivalent to Theorem 3.42 is the following:

Theorem 3.43. Let S be a finite set. Let G be a finite set. For each $i \in G$, let A_i be a subset of S. We define the intersection $\bigcap_{i\in\varnothing} A_i$ (which would otherwise be undefined, since $\varnothing$ is the empty set) to mean the set S. (Thus, $\bigcap_{i\in I} A_i$ is defined for any subset I of G, not just for nonempty subsets I.) Then,

$$\left|S\setminus\left(\bigcup_{i\in G} A_i\right)\right| = \sum_{I\subseteq G} (-1)^{|I|}\left|\bigcap_{i\in I} A_i\right|.$$

Theorem 3.42 and Theorem 3.43 are commonly known as the *principle of inclusion and exclusion*, or as the *Sylvester sieve formula*. They are not hard to prove (see, e.g., [Galvin17, §16] for two proofs). Rather than proving them directly, we shall however generalize them and prove the generalization, from which they are easily obtained as particular cases.

We generalize these theorems in two steps. First, we observe that the $S\setminus\left(\bigcup_{i\in G} A_i\right)$ in Theorem 3.43 is simply the set of all elements of S that belong to none of the subsets A_i (for $i \in G$). In other words,

$$S\setminus\left(\bigcup_{i\in G} A_i\right) = \{s \in S \mid \text{the number of } i \in G \text{ satisfying } s \in A_i \text{ equals } 0\}.$$

We can replace the "0" here by any number k, and ask for the size of the resulting set. The answer is given by the following result of Charles Jordan (see [Comtet74, §4.8, Theorem A] and [DanRot78] for fairly complicated proofs):

Theorem 3.44. Let S be a finite set. Let G be a finite set. For each $i \in G$, let A_i be a subset of S. We define the intersection $\bigcap_{i \in \varnothing} A_i$ (which would otherwise be undefined, since $\varnothing$ is the empty set) to mean the set S. (Thus, $\bigcap_{i \in I} A_i$ is defined for any subset I of G, not just for nonempty subsets I.)

Let $k \in \mathbb{N}$. Let

$$S_k = \{s \in S \mid \text{ the number of } i \in G \text{ satisfying } s \in A_i \text{ equals } k\}.$$

(In other words, S_k is the set of all elements of S that belong to exactly k of the subsets A_i.) Then,

$$|S_k| = \sum_{I \subseteq G} (-1)^{|I|-k} \binom{|I|}{k} \left| \bigcap_{i \in I} A_i \right|.$$

A different generalization of Theorem 3.43 (closely related to the *Bonferroni inequalities*, for which see [Galvin17, §17, problem (2)]) explores what happens when the sum on the right hand side of the formula is restricted to only those subsets I of G whose size doesn't surpass a given integer m:

Theorem 3.45. Let S be a finite set. Let G be a finite set. For each $i \in G$, let A_i be a subset of S. We define the intersection $\bigcap_{i \in \varnothing} A_i$ (which would otherwise be undefined, since $\varnothing$ is the empty set) to mean the set S. (Thus, $\bigcap_{i \in I} A_i$ is defined for any subset I of G, not just for nonempty subsets I.)

Let $m \in \mathbb{N}$. For each $s \in S$, let $c(s)$ denote the number of $i \in G$ satisfying $s \in A_i$. Then,

$$(-1)^m \sum_{s \in S} \binom{c(s) - 1}{m} = \sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right|.$$

Finally, Theorem 3.44 and Theorem 3.45 can be merged into one common general principle:

Theorem 3.46. Let S be a finite set. Let G be a finite set. For each $i \in G$, let A_i be a subset of S. We define the intersection $\bigcap_{i \in \varnothing} A_i$ (which would otherwise be undefined, since $\varnothing$ is the empty set) to mean the set S. (Thus, $\bigcap_{i \in I} A_i$ is defined for any subset I of G, not just for nonempty subsets I.)

Let $k \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that $m \geq k$. For each $s \in S$, let $c(s)$ denote the number of $i \in G$ satisfying $s \in A_i$. Then,

$$(-1)^m \sum_{s \in S} \binom{c(s)}{k} \binom{c(s)-k-1}{m-k} = \sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \binom{|I|}{k} \left| \bigcap_{i \in I} A_i \right|.$$

As we said, we shall first prove Theorem 3.46, and then derive all the preceding theorems in this section from it. The proof of Theorem 3.46 will rely on several ingredients, the first of which is the following simple identity:

Lemma 3.47. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Let $m \in \{k, k+1, k+2, \ldots\}$. Then,

$$\sum_{r=0}^{m} (-1)^r \binom{n}{r} \binom{r}{k} = (-1)^m \binom{n}{k} \binom{n-k-1}{m-k}.$$

Exercise 3.11. Prove Lemma 3.47.

Next, we introduce a simple yet immensely helpful notation that will facilitate our proof:

Definition 3.48. If $\mathcal{A}$ is any logical statement, then we define an integer $[\mathcal{A}] \in \{0, 1\}$ by

$$[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases}.$$

For example, $[1+1=2] = 1$ (since $1+1=2$ is true), whereas $[1+1=1] = 0$ (since $1+1=1$ is false).

If $\mathcal{A}$ is any logical statement, then the integer $[\mathcal{A}]$ is known as the *truth value* of $\mathcal{A}$. The notation $[\mathcal{A}]$ is known as the *Iverson bracket notation*.

Clearly, if $\mathcal{A}$ and $\mathcal{B}$ are two equivalent logical statements, then $[\mathcal{A}] = [\mathcal{B}]$. This and a few other useful properties of the Iverson bracket notation are collected in the following exercise:

Exercise 3.12. Prove the following rules for truth values:

(a) If $\mathcal{A}$ and $\mathcal{B}$ are two equivalent logical statements, then $[\mathcal{A}] = [\mathcal{B}]$.

(b) If $\mathcal{A}$ is any logical statement, then $[\text{not } \mathcal{A}] = 1 - [\mathcal{A}]$.

(c) If $\mathcal{A}$ and $\mathcal{B}$ are two logical statements, then $[\mathcal{A} \wedge \mathcal{B}] = [\mathcal{A}] [\mathcal{B}]$.

(d) If $\mathcal{A}$ and $\mathcal{B}$ are two logical statements, then $[\mathcal{A} \vee \mathcal{B}] = [\mathcal{A}] + [\mathcal{B}] - [\mathcal{A}] [\mathcal{B}]$.

(e) If $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are three logical statements, then

$$[\mathcal{A} \vee \mathcal{B} \vee \mathcal{C}] = [\mathcal{A}] + [\mathcal{B}] + [\mathcal{C}] - [\mathcal{A}] [\mathcal{B}] - [\mathcal{A}] [\mathcal{C}] - [\mathcal{B}] [\mathcal{C}] + [\mathcal{A}] [\mathcal{B}] [\mathcal{C}].$$

The Iverson bracket helps us rewrite the cardinality of a set as a sum:

Lemma 3.49. Let S be a finite set. Let T be a subset of S. Then,

$$|T| = \sum_{s \in S} [s \in T].$$

Proof of Lemma 3.49. We have

$$\begin{aligned}
\sum_{s \in S} [s \in T] &= \underbrace{\sum_{\substack{s \in S; \\ s \in T}}}_{\substack{= \sum_{s \in T} \\ \text{(since } T \text{ is a} \\ \text{subset of } S)}} \underbrace{[s \in T]}_{\substack{=1 \\ \text{(since } s \in T \text{ is true)}}} + \sum_{\substack{s \in S; \\ s \notin T}} \underbrace{[s \in T]}_{\substack{=0 \\ \text{(since } s \in T \text{ is false} \\ \text{(since } s \notin T))}} \\
&\qquad \text{(since each } s \in S \text{ satisfies either } s \in T \text{ or } s \notin T \text{ (but not both))} \\
&= \underbrace{\sum_{s \in T} 1}_{= |T| \cdot 1} + \underbrace{\sum_{\substack{s \in S; \\ s \notin T}} 0}_{=0} = |T| \cdot 1 + 0 = |T|.
\end{aligned}$$

This proves Lemma 3.49. □

We can now state the main precursor to Theorem 3.46:

Lemma 3.50. Let S be a finite set. Let G be a finite set. For each $i \in G$, let A_i be a subset of S. We define the intersection $\bigcap_{i \in \varnothing} A_i$ (which would otherwise be undefined, since $\varnothing$ is the empty set) to mean the set S. (Thus, $\bigcap_{i \in I} A_i$ is defined for any subset I of G, not just for nonempty subsets I.)

Let $k \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that $m \geq k$.

Let $s \in S$. Let $c(s)$ denote the number of $i \in G$ satisfying $s \in A_i$. Then,

$$\sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \binom{|I|}{k} \left[s \in \bigcap_{i \in I} A_i \right] = (-1)^m \binom{c(s)}{k} \binom{c(s) - k - 1}{m - k}.$$

Proof of Lemma 3.50. From $m \geq k$ and $m \in \mathbb{N}$, we obtain $m \in \{k, k+1, k+2, \ldots\}$.

Define a subset C of G by

$$C = \{i \in G \mid s \in A_i\}.$$

Thus,

$$\begin{aligned}
|C| &= |\{i \in G \mid s \in A_i\}| \\
&= (\text{the number of } i \in G \text{ satisfying } s \in A_i) = c(s)
\end{aligned}$$

(since $c(s)$ was defined as the number of $i \in G$ satisfying $s \in A_i$). In other words, C is a $c(s)$-element set. Hence, for each $r \in \mathbb{N}$, Proposition 3.12 (applied to $c(s)$, r and C instead of m, n and S) shows that

$$\dbinom{c(s)}{r} \text{ is the number of all } r\text{-element subsets of } C. \tag{283}$$

Let I be a subset of G. We have the following equivalence:

$$\left(s \in \bigcap_{i \in I} A_i \right) \iff (s \in A_i \text{ for all } i \in I) \tag{284}$$

[130].

But if $i \in I$, then we have the equivalence

$$(i \in C) \iff (s \in A_i) \tag{285}$$

(since $C = \{i \in G \mid s \in A_i\}$).

Hence, the equivalence (284) becomes

$$\left(s \in \bigcap_{i \in I} A_i \right) \iff \left(\underbrace{s \in A_i}_{\substack{\iff (i \in C) \\ \text{(by (285))}}} \text{ for all } i \in I \right) \iff (i \in C \text{ for all } i \in I) \iff (I \subseteq C).$$

In other words, the two statements $\left(s \in \bigcap\limits_{i \in I} A_i \right)$ and $(I \subseteq C)$ are equivalent. Hence, Exercise 3.12 **(a)** (applied to $\mathcal{A} = \left(s \in \bigcap\limits_{i \in I} A_i \right)$ and $\mathcal{B} = (I \subseteq C)$) shows that

$$\left[s \in \bigcap_{i \in I} A_i \right] = [I \subseteq C]. \tag{286}$$

[130] *Proof of (284):* If I is nonempty, then the equivalence (284) follows immediately from the equality

$$\bigcap_{i \in I} A_i = \{x \mid x \in A_i \text{ for all } i \in I\}$$

(which is the definition of the intersection $\bigcap\limits_{i \in I} A_i$). Thus, for the rest of this proof, we WLOG assume that I is not nonempty.

Hence, the set I is empty. In other words, $I = \varnothing$. Hence, $\bigcap\limits_{i \in I} A_i = \bigcap\limits_{i \in \varnothing} A_i = S$. Thus, $s \in S = \bigcap\limits_{i \in I} A_i$. Hence, the statement $\left(s \in \bigcap\limits_{i \in I} A_i \right)$ is true.

Also, there exist no $i \in I$ (since the set I is empty). Hence, the statement $(s \in A_i \text{ for all } i \in I)$ is vacuously true.

Thus, the statements $\left(s \in \bigcap\limits_{i \in I} A_i \right)$ and $(s \in A_i \text{ for all } i \in I)$ are both true, and therefore equivalent. This proves the equivalence (284).

Now, forget that we fixed I. We thus have proven the equality (286) for every subset I of G.

Now,

$$\sum_{\substack{I\subseteq G;\\ |I|\leq m}} (-1)^{|I|} \binom{|I|}{k} \underbrace{\left[s\in\bigcap_{i\in I} A_i\right]}_{\substack{=[I\subseteq C]\\ \text{(by (286))}}}$$

$$= \sum_{\substack{I\subseteq G;\\ |I|\leq m}} (-1)^{|I|} \binom{|I|}{k} [I\subseteq C]$$

$$= \underbrace{\sum_{\substack{I\subseteq G;\\ |I|\leq m;\\ I\subseteq C}}}_{\substack{=\sum_{\substack{I\subseteq G;\\ I\subseteq C;\\ |I|\leq m}} = \sum_{\substack{I\subseteq C;\\ |I|\leq m}}\\ \text{(since } C \text{ is a subset of } G)}} (-1)^{|I|} \binom{|I|}{k} \underbrace{[I\subseteq C]}_{\substack{=1\\ (\text{since } I\subseteq C)}} + \sum_{\substack{I\subseteq G;\\ |I|\leq m;\\ \text{not } I\subseteq C}} (-1)^{|I|} \binom{|I|}{k} \underbrace{[I\subseteq C]}_{\substack{=0\\ (\text{since we don't have } I\subseteq C)}}$$

$$\left(\begin{array}{c}\text{since each subset } I \text{ of } G \text{ satisfies either } I\subseteq C\\ \text{or } (\text{not } I\subseteq C) \text{ (but not both)}\end{array}\right)$$

$$= \sum_{\substack{I\subseteq C;\\ |I|\leq m}} (-1)^{|I|} \binom{|I|}{k} + \underbrace{\sum_{\substack{I\subseteq G;\\ |I|\leq m;\\ \text{not } I\subseteq C}} (-1)^{|I|} \binom{|I|}{k} 0}_{=0} = \underbrace{\sum_{\substack{I\subseteq C;\\ |I|\leq m}}}_{=\sum_{r=0}^{m} \sum_{\substack{I\subseteq C;\\ |I|=r}}} (-1)^{|I|} \binom{|I|}{k}$$

$$= \sum_{r=0}^{m} \sum_{\substack{I\subseteq C;\\ |I|=r}} \underbrace{(-1)^{|I|}}_{\substack{=(-1)^r\\ (\text{since } |I|=r)}} \underbrace{\binom{|I|}{k}}_{\substack{=\binom{r}{k}\\ (\text{since } |I|=r)}}$$

$$= \sum_{r=0}^{m} \underbrace{\sum_{\substack{I \subseteq C; \\ |I| = r}} (-1)^r \binom{r}{k}}_{= (\text{the number of all subsets } I \text{ of } C \text{ satisfying } |I|=r)(-1)^r \binom{r}{k}}$$

$$= \sum_{r=0}^{m} \underbrace{(\text{the number of all subsets } I \text{ of } C \text{ satisfying } |I| = r)}_{\substack{= (\text{the number of all } r\text{-element subsets of } C) = \binom{c(s)}{r} \\ (\text{by (283)})}} (-1)^r \binom{r}{k}$$

$$= \sum_{r=0}^{m} \binom{c(s)}{r} (-1)^r \binom{r}{k} = \sum_{r=0}^{m} (-1)^r \binom{c(s)}{r} \binom{r}{k} = (-1)^m \binom{c(s)}{k} \binom{c(s) - k - 1}{m - k}$$
$$(\text{by Lemma 3.47 (applied to } n = c(s))).$$

This proves Lemma 3.50. □

We now easily obtain Theorem 3.46:

Proof of Theorem 3.46. For each subset I of G, the intersection $\bigcap_{i \in I} A_i$ is a subset of S [131]. Hence, for each subset I of G, we obtain

$$\left| \bigcap_{i \in I} A_i \right| = \sum_{s \in S} \left[s \in \bigcap_{i \in I} A_i \right] \qquad (287)$$

[131] *Proof.* Let I be a subset of G. We must show that the intersection $\bigcap_{i \in I} A_i$ is a subset of S.

If I is nonempty, then this is clear (because each of the sets A_i is a subset of S). Hence, for the rest of this proof, we can WLOG assume that I is not nonempty. Assume this.

The set I is empty (since I is not nonempty). Hence, $I = \varnothing$. Thus, $\bigcap_{i \in I} A_i = \bigcap_{i \in \varnothing} A_i = S$ (since we defined $\bigcap_{i \in \varnothing} A_i$ to be S). Hence, $\bigcap_{i \in \varnothing} A_i$ is a subset of S. Qed.

(by Lemma 3.49 (applied to $T = \bigcap_{i \in I} A_i$)). Hence,

$$\sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \dbinom{|I|}{k} \underbrace{\left| \bigcap_{i \in I} A_i \right|}_{\substack{= \sum_{s \in S} \left[s \in \bigcap_{i \in I} A_i \right] \\ \text{(by (287))}}}$$

$$= \sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \dbinom{|I|}{k} \sum_{s \in S} \left[s \in \bigcap_{i \in I} A_i \right] = \underbrace{\sum_{\substack{I \subseteq G; \\ |I| \leq m}} \sum_{s \in S}}_{= \sum_{s \in S} \sum_{\substack{I \subseteq G; \\ |I| \leq m}}} (-1)^{|I|} \dbinom{|I|}{k} \left[s \in \bigcap_{i \in I} A_i \right]$$

$$= \sum_{s \in S} \underbrace{\sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \dbinom{|I|}{k} \left[s \in \bigcap_{i \in I} A_i \right]}_{\substack{= (-1)^m \dbinom{c(s)}{k} \dbinom{c(s) - k - 1}{m - k} \\ \text{(by Lemma 3.50)}}}$$

$$= \sum_{s \in S} (-1)^m \dbinom{c(s)}{k} \dbinom{c(s) - k - 1}{m - k} = (-1)^m \sum_{s \in S} \dbinom{c(s)}{k} \dbinom{c(s) - k - 1}{m - k}.$$

This proves Theorem 3.46. □

Having proven Theorem 3.46, we can now easily derive the other (less general) versions of the inclusion-exclusion principle:

Exercise 3.13. Prove Theorem 3.42, Theorem 3.43, Theorem 3.44 and Theorem 3.45.

3.6. Additional exercises

This section contains some further exercises. These will not be used in the rest of the notes, and they can be skipped at will[132]. I provide solutions to only a few of them.

Exercise 3.14. Find a different proof of Proposition 3.32 **(f)** that uses a double-counting argument (i.e., counting some combinatorial objects in two different ways, and then concluding that the results are equal).

[Hint: How many $(x + y + 1)$-element subsets does the set $\{1, 2, \ldots, n + 1\}$ have? Now, for a given $k \in \{0, 1, \ldots, n\}$, how many $(x + y + 1)$-element subsets whose $(x + 1)$-th smallest element is $k + 1$ does the set $\{1, 2, \ldots, n + 1\}$ have?]

[132] The same, of course, can be said for many of the standard exercises.

Exercise 3.15. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ be fixed. Show that the number of all k-tuples $(a_1, a_2, \ldots, a_k) \in \mathbb{N}^k$ satisfying $a_1 + a_2 + \cdots + a_k = n$ equals $\binom{n+k-1}{n}$.

Remark 3.51. Exercise 3.15 can be restated in terms of multisets. Namely, let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ be fixed. Also, fix a k-element set K. Then, the number of n-element multisets whose elements all belong to K is $\binom{n+k-1}{n}$. Indeed, we can WLOG assume that $K = \{1, 2, \ldots, k\}$ (otherwise, just relabel the elements of K); then, the multisets whose elements all belong to K are in bijection with the k-tuples $(a_1, a_2, \ldots, a_k) \in \mathbb{N}^k$. The bijection sends a multiset M to the k-tuple $(m_1(M), m_2(M), \ldots, m_k(M))$, where each $m_i(M)$ is the multiplicity of the element i in M. The size of a multiset M corresponds to the sum $a_1 + a_2 + \cdots + a_k$ of the entries of the resulting k-tuple; thus, we get a bijection between

- the n-element multisets whose elements all belong to K

and

- the k-tuples $(a_1, a_2, \ldots, a_k) \in \mathbb{N}^k$ satisfying $a_1 + a_2 + \cdots + a_k = n$.

As a consequence, Exercise 3.15 shows that the number of the former multisets is $\binom{n+k-1}{n}$.

Similarly, we can reinterpret the classical combinatorial interpretation of $\binom{k}{n}$ (as the number of n-element subsets of $\{1, 2, \ldots, k\}$) as follows: The number of all k-tuples $(a_1, a_2, \ldots, a_k) \in \{0, 1\}^k$ satisfying $a_1 + a_2 + \cdots + a_k = n$ equals $\binom{k}{n}$.

See [Galvin17, §13] and [Loehr11, §1.11] for more about multisets.

Exercise 3.16. Let n and a be two integers with $n \geq a \geq 1$. Prove that

$$\sum_{k=a}^{n} \frac{(-1)^k}{k} \binom{n-a}{k-a} = \frac{(-1)^a}{a \binom{n}{a}}.$$

Exercise 3.17. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Prove that

$$\sum_{u=0}^{k} \binom{n+u-1}{u} \binom{n}{k-2u} = \binom{n+k-1}{k}.$$

Here, $\binom{a}{b}$ is defined to be 0 when $b < 0$.

Exercise 3.17 is solved in [Grinbe16a].

Exercise 3.18. Let $N \in \mathbb{N}$. The *binomial transform* of a finite sequence $(f_0, f_1, \ldots, f_N) \in \mathbb{Z}^{N+1}$ is defined to be the sequence $(g_0, g_1, \ldots, g_N)$ defined by

$$g_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} f_i \qquad \text{for every } n \in \{0, 1, \ldots, N\}.$$

(a) Let $(f_0, f_1, \ldots, f_N) \in \mathbb{Z}^{N+1}$ be a finite sequence of integers. Let $(g_0, g_1, \ldots, g_N)$ be the binomial transform of $(f_0, f_1, \ldots, f_N)$. Show that $(f_0, f_1, \ldots, f_N)$ is, in turn, the binomial transform of $(g_0, g_1, \ldots, g_N)$.

(b) Find the binomial transform of the sequence $(1, 1, \ldots, 1)$.

(c) For any given $a \in \mathbb{N}$, find the binomial transform of the sequence $\left(\binom{0}{a}, \binom{1}{a}, \ldots, \binom{N}{a}\right)$.

(d) For any given $q \in \mathbb{Z}$, find the binomial transform of the sequence $(q^0, q^1, \ldots, q^N)$.

(e) Find the binomial transform of the sequence $(1, 0, 1, 0, 1, 0, \ldots)$ (this ends with 1 if N is even, and with 0 if N is odd).

(f) Let $B : \mathbb{Z}^{N+1} \to \mathbb{Z}^{N+1}$ be the map which sends every sequence $(f_0, f_1, \ldots, f_N) \in \mathbb{Z}^{N+1}$ to its binomial transform $(g_0, g_1, \ldots, g_N) \in \mathbb{Z}^{N+1}$. Thus, part **(a)** of this exercise states that $B^2 = \operatorname{id}$.

On the other hand, let $W : \mathbb{Z}^{N+1} \to \mathbb{Z}^{N+1}$ be the map which sends every sequence $(f_0, f_1, \ldots, f_N) \in \mathbb{Z}^{N+1}$ to $\left((-1)^N f_N, (-1)^N f_{N-1}, \ldots, (-1)^N f_0\right) \in \mathbb{Z}^{N+1}$. It is rather clear that $W^2 = \operatorname{id}$.

Show that, furthermore, $B \circ W \circ B = W \circ B \circ W$ and $(B \circ W)^3 = \operatorname{id}$.

Exercise 3.19. Let $n \in \mathbb{N}$. Prove that

$$\sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}.$$

[**Hint:** How does the left hand side grow when n is replaced by $n+1$?]

Exercise 3.19 is taken from [AndFen04, Example 3.7].

Exercise 3.20. Let $n \in \mathbb{N}$.

(a) Prove that

$$\sum_{k=0}^{n} \frac{(-1)^k}{\binom{n}{k}} = 2 \cdot \frac{n+1}{n+2} [n \text{ is even}].$$

(Here, we are using the Iverson bracket notation, as in Definition 3.48; thus, $[n \text{ is even}]$ is 1 if n is even and 0 otherwise.)

(b) Prove that

$$\sum_{k=0}^{n} \frac{1}{\binom{n}{k}} = \frac{n+1}{2^{n+1}} \sum_{k=1}^{n+1} \frac{2^k}{k}.$$

[**Hint:** Show that $\frac{1}{\binom{n}{k}} = \left(\frac{1}{\binom{n+1}{k}} + \frac{1}{\binom{n+1}{k+1}} \right) \frac{n+1}{n+2}$ for each $k \in \{0, 1, \ldots, n\}$.]

Exercise 3.20 **(a)** is [KurLis78, (8)]. Exercise 3.20 **(b)** is [KurLis78, (9)] and [AndFen04, Example 3.9] and [AndDos12, Lemma 3.14] and part of [Rocket81, Theorem 1], and also appears with proof in `https://math.stackexchange.com/a/481686/` (where it is used to show that $\lim_{n\to\infty} \sum_{k=0}^{n} \frac{1}{\binom{n}{k}} = 2$).

Exercise 3.21. For any $n \in \mathbb{N}$ and $m \in \mathbb{N}$, define a polynomial $Z_{m,n} \in \mathbb{Z}[X]$ by

$$Z_{m,n} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(X^{n-k} - 1 \right)^m.$$

Show that $Z_{m,n} = Z_{n,m}$ for any $n \in \mathbb{N}$ and $m \in \mathbb{N}$.

Exercise 3.22. Let $n \in \mathbb{N}$. Prove

$$\sum_{k=0}^{n} (-1)^k \binom{X}{k} \binom{X}{n-k} = \begin{cases} (-1)^{n/2} \binom{X}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

(an identity between polynomials in $\mathbb{Q}[X]$).

[**Hint:** It is enough to prove this when X is replaced by a nonnegative integer r (why?). Now that you have gotten rid of polynomials, introduce new polynomials. Namely, compute the coefficient of X^n in $(1+X)^r (1-X)^r$. Compare with the coefficient of X^n in $\left(1 - X^2\right)^r$.]

Exercise 3.23. Let $n \in \mathbb{N}$.

(a) Prove that

$$\sum_{k=0}^{n} \binom{2k}{k} \binom{2(n-k)}{n-k} = 4^n.$$

(b) Prove that

$$\sum_{k=0}^{n} (-1)^k \binom{2k}{k} \binom{2(n-k)}{n-k} = \begin{cases} 2^n \binom{n}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

[**Hint:** Recall Exercise 3.2 **(b)**.]

Exercise 3.24. Let m be a positive integer. Prove the following:

(a) The binomial coefficient $\dbinom{2m}{m}$ is even.

(b) If m is odd and satisfies $m > 1$, then the binomial coefficient $\dbinom{2m-1}{m-1}$ is even.

(c) If m is odd and satisfies $m > 1$, then $\dbinom{2m}{m} \equiv 0 \bmod 4$.

Exercise 3.25. For any $m \in \mathbb{N}$ and $n \in \mathbb{N}$, define a rational number $T(m,n)$ by

$$T(m,n) = \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}.$$

Prove the following facts:

(a) We have $4T(m,n) = T(m+1,n) + T(m,n+1)$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(b) We have $T(m,n) \in \mathbb{N}$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(c) If $m \in \mathbb{N}$ and $n \in \mathbb{N}$ are such that $(m,n) \neq (0,0)$, then the integer $T(m,n)$ is even.

(d) If $m \in \mathbb{N}$ and $n \in \mathbb{N}$ are such that $m+n$ is odd and $m+n > 1$, then $4 \mid T(m,n)$.

(e) We have $T(m,0) = \dbinom{2m}{m}$ for every $m \in \mathbb{N}$.

(f) We have $T(m,n) = \dfrac{\dbinom{2m}{m}\dbinom{2n}{n}}{\dbinom{m+n}{m}}$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(g) We have $T(m,n) = T(n,m)$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(h) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m,n\}$. Then,

$$\sum_{k=-p}^{p} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \binom{m+n}{m}.$$

(i) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m,n\}$. Then,

$$T(m,n) = \sum_{k=-p}^{p} (-1)^k \binom{2m}{m+k}\binom{2n}{n-k}.$$

Remark 3.52. The numbers $T(m,n)$ introduced in Exercise 3.25 are the so-called *super-Catalan numbers*; much has been written about them (e.g., [Gessel92] and [AleGhe14]). Exercise 3.25 **(b)** suggests that these numbers count something, but no one has so far discovered what. Exercise 3.25 **(i)** is a result of von Szily (1894); see [Gessel92, (29)]. Exercise 3.25 **(b)** is a result of Eugène Catalan (1874), and has also been posed as Problem 3 of the International Mathematical Olympiad 1972. Parts of Exercise 3.25 are also discussed on the thread `https://artofproblemsolving.com/community/c6h1553916s1_supercatalan_numbers` .

The following exercise is a variation on (238):

Exercise 3.26. Let a and b be two integers such that $b \neq 0$. Let $n \in \mathbb{N}$. Show that there exists some $N \in \mathbb{N}$ such that $b^N \dbinom{a/b}{n} \in \mathbb{Z}$.

[**Hint:** I am not aware of a combinatorial solution to this exercise! (I.e., I don't know what the numbers $b^N \dbinom{a/b}{n}$ count, even when they are nonnegative.) All solutions that I know use some (elementary) number theory. For the probably slickest (although unmotivated) solution, basic modular arithmetic suffices; here is a roadmap: First, show that if b and c are integers such that $c > 0$, then there exists an $s \in \mathbb{Z}$ such that $b^{c-1} \equiv sb^c \mod c$ [133]. Apply this to $c = n!$ and conclude that $b^{n!}(a/b - i) \equiv b^{n!}(sa - i) \mod n!$ for every $i \in \mathbb{Z}$. Now use $\dbinom{sa}{n} \in \mathbb{Z}$.]

Exercise 3.27. **(a)** If x and y are two real numbers such that $x + y = 1$, and if $n \in \mathbb{N}$ and $m \in \mathbb{N}$, then prove that

$$x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k + y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k = 1.$$

(b) Let $n \in \mathbb{N}$. Prove that

$$\sum_{k=0}^{n} \binom{n+k}{k} \frac{1}{2^k} = 2^n.$$

Remark 3.53. Exercise 3.27 **(a)** is Problem 7 from the IMO Shortlist 1975. It is also closely related to the Daubechies identity [Zeilbe93] (indeed, the first equality in [Zeilbe93] follows by applying it to p, $1-p$, $n-1$ and $n-1$ instead of n and m). Exercise 3.27 **(b)** is a fairly well-known identity for binomial coefficients (see, e.g., [GrKnPa94, (5.20)]).

[133] To prove this, argue that at least two of $b^0, b^1, \ldots, b^c$ are congruent modulo c.

4. Recurrent sequences

4.1. Basics

Two of the most famous integer sequences defined recursively are the Fibonacci sequence and the Lucas sequence:

- The *Fibonacci sequence* is the sequence $(f_0, f_1, f_2, \ldots)$ of integers which is defined recursively by $f_0 = 0$, $f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 2$. We have already introduced this sequence in Example 2.25. Its first terms are

  $$f_0 = 0, \quad f_1 = 1, \quad f_2 = 1, \quad f_3 = 2, \quad f_4 = 3, \quad f_5 = 5,$$
  $$f_6 = 8, \quad f_7 = 13, \quad f_8 = 21, \quad f_9 = 34, \quad f_{10} = 55,$$
  $$f_{11} = 89, \quad f_{12} = 144, \quad f_{13} = 233.$$

  (Some authors[134] prefer to start the sequence at f_1 rather than f_0; of course, the recursive definition then needs to be modified to require $f_2 = 1$ instead of $f_0 = 0$.)

- The *Lucas sequence* is the sequence $(\ell_0, \ell_1, \ell_2, \ldots)$ of integers which is defined recursively by $\ell_0 = 2$, $\ell_1 = 1$, and $\ell_n = \ell_{n-1} + \ell_{n-2}$ for all $n \geq 2$. Its first terms are

  $$\ell_0 = 2, \quad \ell_1 = 1, \quad \ell_2 = 3, \quad \ell_3 = 4, \quad \ell_4 = 7, \quad \ell_5 = 11,$$
  $$\ell_6 = 18, \quad \ell_7 = 29, \quad \ell_8 = 47, \quad \ell_9 = 76, \quad \ell_{10} = 123,$$
  $$\ell_{11} = 199, \quad \ell_{12} = 322, \quad \ell_{13} = 521.$$

A lot of papers and even books have been written about these two sequences, the relations between them, and the identities that hold for their terms.[135] One of their most striking properties is that they can be computed explicitly, albeit using irrational numbers. In fact, the *Binet formula* says that the n-th Fibonacci number f_n can be computed by

$$f_n = \frac{1}{\sqrt{5}} \varphi^n - \frac{1}{\sqrt{5}} \psi^n, \qquad (288)$$

where $\varphi = \dfrac{1+\sqrt{5}}{2}$ and $\psi = \dfrac{1-\sqrt{5}}{2}$ are the two solutions of the quadratic equation $X^2 - X - 1 = 0$. (The number φ is known as the *golden ratio*; the number ψ can be obtained from it by $\psi = 1 - \varphi = -1/\varphi$.) A similar formula, using the very same numbers φ and ψ, exists for the Lucas numbers:

$$\ell_n = \varphi^n + \psi^n. \qquad (289)$$

---

[134] such as Vorobiev in his book [Vorobi02]

[135] See `https://oeis.org/A000045` and `https://oeis.org/A000032` for an overview of their properties. The book [Vorobi02] is a readable introduction to the Fibonacci sequence, which also surveys a lot of other mathematics (elementary number theory, continued fractions, and even some geometry) along the way. Another introduction to the Fibonacci sequence is [CamFon07].

Remark 4.1. How easy is it to compute f_n and ℓ_n using the formulas (288) and (289)?

This is a nontrivial question. Indeed, if you are careless, you may find them rather useless. For instance, if you try to compute f_n using the formula (288) and using approximate values for the irrational numbers φ and ψ, then you might end up with a wrong value for f_n, because the error in the approximate value for φ propagates when you take φ to the n-th power. (And for high enough n, the error will become larger than 1, so you will not be able to get the correct value by rounding.) The greater n is, the more precise you need a value for φ to approximate f_n this way. Thus, approximating φ is not a good way to compute f_n. (Actually, the opposite is true: You can use (288) to approximate φ by computing Fibonacci numbers. Namely, it is easy to show that $\varphi = \lim\limits_{n\to\infty} \dfrac{f_n}{f_{n-1}}$.)

A better approach to using (288) is to work with the exact values of φ and ψ. To do so, you need to know how to add, subtract, multiply and divide real numbers of the form $a + b\sqrt{5}$ with $a, b \in \mathbb{Q}$ without ever using approximations. (Clearly, φ, ψ and $\sqrt{5}$ all have this form.) There are rules for this, which are simple to check:

$$\begin{aligned}
\left(a+b\sqrt{5}\right)+\left(c+d\sqrt{5}\right) &= (a+c)+(b+d)\sqrt{5};\\
\left(a+b\sqrt{5}\right)-\left(c+d\sqrt{5}\right) &= (a-c)+(b-d)\sqrt{5};\\
\left(a+b\sqrt{5}\right)\cdot\left(c+d\sqrt{5}\right) &= (ac+5bd)+(bc+ad)\sqrt{5};\\
\frac{a+b\sqrt{5}}{c+d\sqrt{5}} &= \frac{(ac-5bd)+(bc-ad)\sqrt{5}}{c^2-5d^2} \qquad \text{for } (c,d)\neq(0,0).
\end{aligned}$$

(The last rule is an instance of "rationalizing the denominator".) These rules give you a way to exactly compute things like φ^n, $\frac{1}{\sqrt{5}}\varphi^n$, ψ^n and $\frac{1}{\sqrt{5}}\psi^n$, and thus also f_n and ℓ_n. If you use exponentiation by squaring to compute n-th powers, this actually becomes a fast algorithm (a lot faster than just computing f_n and ℓ_n using the recurrence). So, yes, (288) and (289) are useful.

We shall now study a generalization of both the Fibonacci and the Lucas sequences, and generalize (288) and (289) to a broader class of sequences.

Definition 4.2. If a and b are two complex numbers, then a sequence $(x_0, x_1, x_2, \ldots)$ of complex numbers will be called (a,b)*-recurrent* if every $n \geq 2$ satisfies

$$x_n = ax_{n-1} + bx_{n-2}.$$

So, the Fibonacci sequence and the Lucas sequence are $(1,1)$-recurrent. An (a,b)-recurrent sequence $(x_0, x_1, x_2, \ldots)$ is fully determined by the four values a, b, x_0

and x_1, and can be constructed for any choice of these four values. Here are some further examples of (a, b)-recurrent sequences:

- The sequence $(x_0, x_1, x_2, \ldots)$ in Theorem 2.26 is (a, b)-recurrent (by its very definition).
- A sequence $(x_0, x_1, x_2, \ldots)$ is $(2, -1)$-recurrent if and only if every $n \geq 2$ satisfies $x_n = 2x_{n-1} - x_{n-2}$. In other words, a sequence $(x_0, x_1, x_2, \ldots)$ is $(2, -1)$-recurrent if and only if every $n \geq 2$ satisfies $x_n - x_{n-1} = x_{n-1} - x_{n-2}$. In other words, a sequence $(x_0, x_1, x_2, \ldots)$ is $(2, -1)$-recurrent if and only if $x_1 - x_0 = x_2 - x_1 = x_3 - x_2 = \cdots$. In other words, the $(2, -1)$-recurrent sequences are precisely the arithmetic progressions.
- Geometric progressions are also (a, b)-recurrent for appropriate a and b. Namely, any geometric progression $\left(u, uq, uq^2, uq^3, \ldots\right)$ is $(q, 0)$-recurrent, since every $n \geq 2$ satisfies $uq^n = q \cdot uq^{n-1} + 0 \cdot uq^{n-2}$. However, not every $(q, 0)$-recurrent sequence $(x_0, x_1, x_2, \ldots)$ is a geometric progression (since the condition $x_n = qx_{n-1} + 0x_{n-2}$ for all $n \geq 2$ says nothing about x_0, and thus x_0 can be arbitrary).
- A sequence $(x_0, x_1, x_2, \ldots)$ is $(0, 1)$-recurrent if and only if every $n \geq 2$ satisfies $x_n = x_{n-2}$. In other words, a sequence $(x_0, x_1, x_2, \ldots)$ is $(0, 1)$-recurrent if and only if it has the form $(u, v, u, v, u, v, \ldots)$ for two complex numbers u and v.
- A sequence $(x_0, x_1, x_2, \ldots)$ is $(1, 0)$-recurrent if and only if every $n \geq 2$ satisfies $x_n = x_{n-1}$. In other words, a sequence $(x_0, x_1, x_2, \ldots)$ is $(1, 0)$-recurrent if and only if it has the form $(u, v, v, v, v, \ldots)$ for two complex numbers u and v. Notice that u is not required to be equal to v, because we never claimed that $x_n = x_{n-1}$ holds for $n = 1$.
- A sequence $(x_0, x_1, x_2, \ldots)$ is $(1, -1)$-recurrent if and only if every $n \geq 2$ satisfies $x_n = x_{n-1} - x_{n-2}$. Curiously, it turns out that every such sequence is 6-periodic (i.e., it satisfies $x_{n+6} = x_n$ for every $n \in \mathbb{N}$), because every $n \in \mathbb{N}$ satisfies

  $$x_{n+6} = \underbrace{x_{n+5}}_{= x_{n+4} - x_{n+3}} - x_{n+4} = (x_{n+4} - x_{n+3}) - x_{n+4} = - \underbrace{x_{n+3}}_{= x_{n+2} - x_{n+1}}$$
  $$= - \left(\underbrace{x_{n+2}}_{= x_{n+1} - x_n} - x_{n+1} \right) = - (x_{n+1} - x_n - x_{n+1}) = x_n.$$

  More precisely, a sequence $(x_0, x_1, x_2, \ldots)$ is $(1, 0)$-recurrent if and only if it has the form $(u, v, v - u, -u, -v, u - v, \ldots)$ (where the "..." stands for "repeat the preceding 6 values over and over" here) for two complex numbers u and v.

- The above three examples notwithstanding, most (a, b)-recurrent sequences of course are not periodic. However, here is another example which provides a great supply of non-periodic (a, b)-recurrent sequences and, at the same time, explains why we get so many periodic ones: If α is any angle, then the sequences

$$(\sin(0\alpha), \sin(1\alpha), \sin(2\alpha), \ldots) \qquad \text{and}$$
$$(\cos(0\alpha), \cos(1\alpha), \cos(2\alpha), \ldots)$$

are $(2\cos\alpha, -1)$-recurrent. More generally, if α and β are two angles, then the sequence

$$(\sin(\beta + 0\alpha), \sin(\beta + 1\alpha), \sin(\beta + 2\alpha), \ldots)$$

is $(2\cos\alpha, -1)$-recurrent[136]. When $\alpha \in 2\pi\mathbb{Q}$ (that is, $\alpha = 2\pi r$ for some $r \in \mathbb{Q}$), this sequence is periodic.

4.2. Explicit formulas (à la Binet)

Now, we can get an explicit formula (similar to (288) and (289)) for every term of an (a, b)-recurrent sequence (in terms of a, b, x_0 and x_1) in the case when $a^2 + 4b \neq 0$. Here is how this works:

Remark 4.3. Let a and b be complex numbers such that $a^2 + 4b \neq 0$. Let $(x_0, x_1, x_2, \ldots)$ be an (a, b)-recurrent sequence. We want to construct an explicit formula for each x_n in terms of x_0, x_1, a and b.

---

[136] *Proof.* Let α and β be two angles. We need to show that the sequence $(\sin(\beta + 0\alpha), \sin(\beta + 1\alpha), \sin(\beta + 2\alpha), \ldots)$ is $(2\cos\alpha, -1)$-recurrent. In other words, we need to prove that

$$\sin(\beta + n\alpha) = 2\cos\alpha \sin(\beta + (n-1)\alpha) + (-1)\sin(\beta + (n-2)\alpha)$$

for every $n \geq 2$. So fix $n \geq 2$.

One of the well-known trigonometric identities states that $\sin x + \sin y = 2\sin\frac{x+y}{2}\cos\frac{x-y}{2}$ for any two angles x and y. Applying this to $x = \beta + n\alpha$ and $y = \beta + (n-2)\alpha$, we obtain

$$\begin{aligned}\sin(\beta + n\alpha) + \sin(\beta + (n-2)\alpha) &= 2\sin\underbrace{\frac{(\beta + n\alpha) + (\beta + (n-2)\alpha)}{2}}_{=\beta+(n-1)\alpha}\cos\underbrace{\frac{(\beta + n\alpha) - (\beta + (n-2)\alpha)}{2}}_{=\alpha} \\ &= 2\sin(\beta + (n-1)\alpha)\cos\alpha = 2\cos\alpha\sin(\beta + (n-1)\alpha).\end{aligned}$$

Hence,

$$\begin{aligned}\sin(\beta + n\alpha) &= 2\cos\alpha\sin(\beta + (n-1)\alpha) - \sin(\beta + (n-2)\alpha) \\ &= 2\cos\alpha\sin(\beta + (n-1)\alpha) + (-1)\sin(\beta + (n-2)\alpha),\end{aligned}$$

qed.

To do so, we let q_+ and q_- be the two solutions of the quadratic equation $X^2 - aX - b = 0$, namely

$$q_+ = \frac{a + \sqrt{a^2 + 4b}}{2} \qquad \text{and} \qquad q_- = \frac{a - \sqrt{a^2 + 4b}}{2}.$$

We notice that $q_+ \neq q_-$ (since $a^2 + 4b \neq 0$). It is easy to see that the sequences $\left(1, q_+, q_+^2, q_+^3, \ldots\right)$ and $\left(1, q_-, q_-^2, q_-^3, \ldots\right)$ are (a, b)-recurrent. As a consequence, for any two complex numbers λ_+ and λ_-, the sequence

$$\left(\lambda_+ + \lambda_-, \lambda_+ q_+ + \lambda_- q_-, \lambda_+ q_+^2 + \lambda_- q_-^2, \ldots\right)$$

(the n-th term of this sequence, with n starting at 0, is $\lambda_+ q_+^n + \lambda_- q_-^n$) must also be (a, b)-recurrent (check this!). We denote this sequence by L_{λ_+, λ_-}.

We now need to find two complex numbers λ_+ and λ_- such that this sequence L_{λ_+, λ_-} is our sequence $(x_0, x_1, x_2, \ldots)$. In order to do so, we only need to ensure that $\lambda_+ + \lambda_- = x_0$ and $\lambda_+ q_+ + \lambda_- q_- = x_1$ (because once this holds, it will follow that the sequences L_{λ_+, λ_-} and $(x_0, x_1, x_2, \ldots)$ have the same first two terms; and this will yield that these two sequences are identical, because two (a, b)-recurrent sequences with the same first two terms must be identical). That is, we need to solve the system of linear equations

$$\begin{cases} \lambda_+ + \lambda_- = x_0; \\ \lambda_+ q_+ + \lambda_- q_- = x_1 \end{cases} \qquad \text{in the unknowns } \lambda_+ \text{ and } \lambda_-.$$

Thanks to $q_+ \neq q_-$, this system has a unique solution:

$$\lambda_+ = \frac{x_1 - q_- x_0}{q_+ - q_-}; \qquad \lambda_- = \frac{q_+ x_0 - x_1}{q_+ - q_-}.$$

Thus, if we set (λ_+, λ_-) to be this solution, then $(x_0, x_1, x_2, \ldots) = L_{\lambda_+, \lambda_-}$, so that

$$x_n = \lambda_+ q_+^n + \lambda_- q_-^n \qquad (290)$$

for every nonnegative integer n. This is an explicit formula, at least if the square roots do not disturb you. When $x_0 = 0$ and $x_1 = a = b = 1$, you get the famous Binet formula (288) for the Fibonacci sequence.

In the next exercise you will see what happens if the $a^2 + 4b \neq 0$ condition does not hold.

Exercise 4.1. Let a and b be complex numbers such that $a^2 + 4b = 0$. Consider an (a, b)-recurrent sequence $(x_0, x_1, x_2, \ldots)$. Find an explicit formula for each x_n in terms of x_0, x_1, a and b.

[**Note:** The polynomial $X^2 - aX - b$ has a double root here. Unlike the case of two distinct roots studied above, you won't see any radicals here. The explicit formula really deserves the name "explicit".]

Remark 4.3 and Exercise 4.1, combined, solve the problem of finding an explicit formula for any term of an (a,b)-recurrent sequence when a and b are complex numbers, at least if you don't mind having square roots in your formula. Similar tactics can be used to find explicit forms for the more general case of sequences satisfying "homogeneous linear recurrences with constant coefficients"[137], although instead of square roots you will now need roots of higher-degree polynomials. (See [LeLeMe16, §22.3.2 ("Solving Homogeneous Linear Recurrences")] for an outline of this; see also [Heffer17, Topic "Linear Recurrences"] for a linear-algebraic introduction.)

4.3. Further results

Here are some more exercises from the theory of recurrent sequences. I am not going particularly deep here, but we may encounter generalizations later.

First, an example: If we "split" the Fibonacci sequence

$$(f_0, f_1, f_2, \ldots) = (0, 1, 1, 2, 3, 5, 8, \ldots)$$

into two subsequences

$$(f_0, f_2, f_4, \ldots) = (0, 1, 3, 8, 21, \ldots) \qquad \text{and} \qquad (f_1, f_3, f_5, \ldots) = (1, 2, 5, 13, \ldots)$$

(each of which contains every other Fibonacci number), then it turns out that each of these two subsequences is $(3, -1)$-recurrent[138]. This is rather easy to prove, but one can always ask for generalizations: What happens if we start with an arbitrary (a,b)-recurrent sequence, instead of the Fibonacci numbers? What happens if we split it into three, four or more subsequences? The answer is rather nice:

Exercise 4.2. Let a and b be complex numbers. Let $(x_0, x_1, x_2, \ldots)$ be an (a,b)-recurrent sequence.

(a) Prove that the sequences $(x_0, x_2, x_4, \ldots)$ and $(x_1, x_3, x_5, \ldots)$ are (c,d)-recurrent for some complex numbers c and d. Find these c and d.

(b) Prove that the sequences $(x_0, x_3, x_6, \ldots)$, $(x_1, x_4, x_7, \ldots)$ and $(x_2, x_5, x_8, \ldots)$ are (c,d)-recurrent for some (other) complex numbers c and d.

(c) For every nonnegative integers N and K, prove that the sequence $(x_K, x_{N+K}, x_{2N+K}, x_{3N+K}, \ldots)$ is (c,d)-recurrent for some complex numbers c and d which depend only on N, a and b (but not on K or x_0 or x_1).

---

[137] These are sequences $(x_0, x_1, x_2, \ldots)$ which satisfy

$$\left(x_n = c_1 x_{n-1} + c_2 x_{n-2} + \cdots + c_k x_{n-k} \qquad \text{for all } n \geq k\right)$$

for a fixed $k \in \mathbb{N}$ and a fixed k-tuple $(c_1, c_2, \ldots, c_k)$ of complex numbers. When $k = 2$, these are the (c_1, c_2)-recurrent sequences.

[138] In other words, we have $f_{2n} = 3 f_{2(n-1)} + (-1) f_{2(n-2)}$ and $f_{2n+1} = 3 f_{2(n-1)+1} + (-1) f_{2(n-2)+1}$ for every $n \geq 2$.

The next exercise gives a combinatorial interpretation of the Fibonacci numbers:

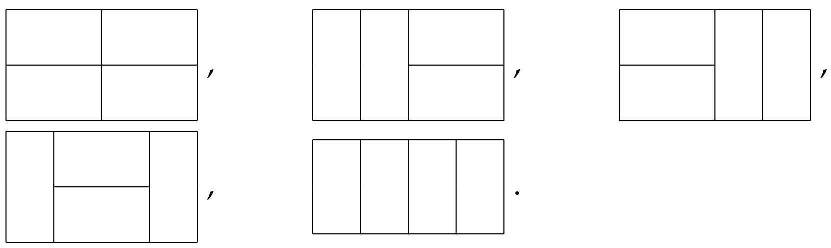
Exercise 4.3. Recall that the Fibonacci numbers $f_0, f_1, f_2, \ldots$ are defined recursively by $f_0 = 0$, $f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 2$. For every positive integer n, show that f_n is the number of subsets I of $\{1, 2, \ldots, n-2\}$ such that no two elements of I are consecutive (i.e., there exists no $i \in \mathbb{Z}$ such that both i and $i+1$ belong to I). For instance, for $n = 5$, these subsets are $\varnothing$, $\{1\}$, $\{2\}$, $\{3\}$ and $\{1, 3\}$.

Notice that $\{1, 2, \ldots, -1\}$ is to be understood as the empty set (since there are no integers x satisfying $1 \leq x \leq -1$). (So Exercise 4.3, applied to $n = 1$, says that f_1 is the number of subsets I of the empty set such that no two elements of I are consecutive. This is correct, because the empty set has only one subset, which of course is empty and thus has no consecutive elements; and the Fibonacci number f_1 is precisely 1.)

Remark 4.4. Exercise 4.3 is equivalent to another known combinatorial interpretation of the Fibonacci numbers.

Namely, let n be a positive integer. Consider a rectangular table of dimensions $2 \times (n-1)$ (that is, with 2 rows and $n-1$ columns). How many ways are there to subdivide this table into dominos? (A *domino* means a set of two adjacent boxes.)

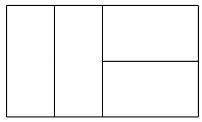
For $n = 5$, there are 5 ways:

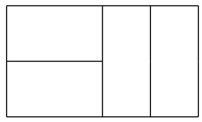


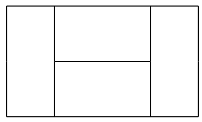
In the general case, there are f_n ways. Why?

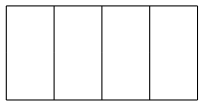
As promised, this result is equivalent to Exercise 4.3. Let us see why. Let P be a way to subdivide the table into dominos. We say that a *horizontal domino* is a domino which consists of two adjacent boxes in the same row; similarly, we define a vertical domino. It is easy to see that (in the subdivision P) each column of the table is covered either by a single vertical domino, or by two horizontal dominos (in which case either both of them "begin" in this column, or both of them "end" in this column). Let $J(P)$ be the set of all $i \in \{1, 2, \ldots, n-1\}$ such that the i-th column of the table is covered by two horizontal dominos, both of which "begin" in this column. For instance,

J () $= \{1, 3\}$;

J () $= \{3\}$;

J () $= \{1\}$;

J () $= \{2\}$;

J () $= \varnothing$.

It is easy to see that the set $J(P)$ is a subset of $\{1, 2, \ldots, n-2\}$ containing no two consecutive integers. Moreover, this set $J(P)$ uniquely determines P, and for every subset I of $\{1, 2, \ldots, n-2\}$ containing no two consecutive integers, there exists some way P to subdivide the table into dominos such that $J(P) = I$.

Hence, the number of all ways to subdivide the table into dominos equals the number of all subsets I of $\{1, 2, \ldots, n-2\}$ containing no two consecutive integers. Exercise 4.3 says that this latter number is f_n; therefore, so is the former number.

(I have made this remark because I found it instructive. If you merely want a proof that the number of all ways to subdivide the table into dominos equals f_n, then I guess it is easier to just prove it by induction without taking the detour through Exercise 4.3. This proof is sketched in [GrKnPa94, §7.1], followed by an informal yet insightful discussion of "infinite sums of dominos" and various related ideas.)

Either Exercise 4.3 or Remark 4.4 can be used to prove properties of Fibonacci numbers in a combinatorial way; see [BenQui04] for some examples of such proofs.

Here is another formula for certain recursive sequences, coming out of a recent paper on cluster algebras[139]:

[139] Specifically, Exercise 4.4 is part of [LeeSch11, Definition 1], but I have reindexed the sequence and fixed the missing upper bound in the sum.

Exercise 4.4. Let $r \in \mathbb{Z}$. Define a sequence $(c_0, c_1, c_2, \ldots)$ of integers recursively by $c_0 = 0$, $c_1 = 1$ and $c_n = rc_{n-1} - c_{n-2}$ for all $n \geq 2$. Show that

$$c_n = \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} r^{n-1-2i} \qquad (291)$$

for every $n \in \mathbb{N}$. Here, we use the following convention: Any expression of the form $a \cdot b$, where a is 0, has to be interpreted as 0, even if b is undefined.[140]

4.4. Additional exercises

This section contains some further exercises. As the earlier "additional exercises", these will not be relied on in the rest of this text, and solutions will not be provided.

Exercise 4.5. Let q and r be two complex numbers. Prove that the sequence $\left(q^0 - r^0, q^1 - r^1, q^2 - r^2, \ldots\right)$ is (a, b)-recurrent for two appropriately chosen a and b. Find these a and b.

Exercise 4.6. Let φ be the golden ratio (i.e., the real number $\dfrac{1+\sqrt{5}}{2}$). Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence.

(a) Show that $f_{n+1} - \varphi f_n = \dfrac{1}{\sqrt{5}} \psi^n$ for every $n \in \mathbb{N}$, where $\psi = \dfrac{1-\sqrt{5}}{2}$. (Notice that $\psi = \dfrac{1-\sqrt{5}}{2} \approx -0.618$ lies between -1 and 0, and thus the powers ψ^n converge to 0 as $n \to \infty$. So $f_{n+1} - \varphi f_n \to 0$ as $n \to \infty$, and consequently $\dfrac{f_{n+1}}{f_n} \to \varphi$ as well.)

(b) Show that

$$f_n = \operatorname{round}\left(\frac{1}{\sqrt{5}} \varphi^n\right) \qquad \text{for every } n \in \mathbb{N}.$$

Here, if x is a real number, then round x denotes the integer closest to x (where, in case of a tie, we take the higher of the two candidates[141]).

[140] The purpose of this convention is to make sure that the right hand side of (291) is well-defined, even though the expression r^{n-1-2i} that appears in it might be undefined (it will be undefined when $r = 0$ and $n - 1 - 2i < 0$).

Of course, the downside of this convention is that we might not have $a \cdot b = b \cdot a$ (because $a \cdot b$ might be well-defined while $b \cdot a$ is not, or vice versa).

[141] This does not really matter in our situation, because $\dfrac{1}{\sqrt{5}} \varphi^n$ will never be a half-integer.

Exercise 4.7. Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence. A set I of integers is said to be *lacunar* if no two elements of I are consecutive (i.e., there exists no $i \in I$ such that $i + 1 \in I$). Show that, for every $n \in \mathbb{N}$, there exists a unique lacunar subset S of $\{2, 3, 4, \ldots\}$ such that $n = \sum_{s \in S} f_s$.

(For example, if $n = 17$, then $S = \{2, 4, 7\}$, because $17 = 1 + 3 + 13 = f_2 + f_4 + f_7$.)

Remark 4.5. The representation of n in the form $n = \sum_{s \in S} f_s$ in Exercise 4.7 is known as the *Zeckendorf representation* of n. It has a number of interesting properties and trivia related to it; for example, there is a rule of thumb for converting miles into kilometers that uses it. It can also be used to define a curious "Fibonacci multiplication" operation on nonnegative integers [Knuth88].

Exercise 4.8. Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence.

(a) Prove the identities

$$\begin{aligned} 1f_n &= f_n && \text{for all } n \geq 0; \\ 2f_n &= f_{n-2} + f_{n+1} && \text{for all } n \geq 2; \\ 3f_n &= f_{n-2} + f_{n+2} && \text{for all } n \geq 2; \\ 4f_n &= f_{n-2} + f_n + f_{n+2} && \text{for all } n \geq 2. \end{aligned}$$

(b) Notice that the right hand sides of these identities have a specific form: they are sums of f_{n+t} for t ranging over a lacunar subset of $\mathbb{Z}$. (See Exercise 4.7 for the definition of "lacunar".) Try to find similar identities for $5f_n$ and $6f_n$.

(c) Prove that such identities exist in general. More precisely, prove the following: Let T be a finite set, and a_t be an integer for every $t \in T$. Then, there exists a unique finite lacunar subset S of $\mathbb{Z}$ such that

$$\sum_{t \in T} f_{n+a_t} = \sum_{s \in S} f_{n+s} \qquad \text{for every } n \in \mathbb{Z} \text{ which}$$
$$\text{satisfies } n \geq \max\left(\{-a_t \mid t \in T\} \cup \{-s \mid s \in S\}\right).$$

(The condition $n \geq \max\left(\{-a_t \mid t \in T\} \cup \{-s \mid s \in S\}\right)$ merely ensures that all the f_{n+a_t} and f_{n+s} are well-defined.)

Remark 4.6. Exercise 4.8 **(c)** is [Grinbe11, Theorem 1.4]. It is also a consequence of [CamFon07, Lemma 6.2] (applied to $k = 2$). I'd be delighted to see other proofs!

Similarly I am highly interested in analogues of Exercises 4.7 and 4.8 for other (a, b)-recurrent sequences (e.g., Lucas numbers).

Exercise 4.9. **(a)** Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence. For every $n \in \mathbb{N}$ and $k \in \mathbb{N}$ satisfying $0 \leq k \leq n$, define a rational number $\dbinom{n}{k}_F$ by

$$\binom{n}{k}_F = \frac{f_n f_{n-1} \cdots f_{n-k+1}}{f_k f_{k-1} \cdots f_1}.$$

This is called the (n,k)-th *Fibonomial coefficient* (in analogy to the binomial coefficient $\dbinom{n}{k} = \dfrac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$).

Show that $\dbinom{n}{k}_F$ is an integer.

(b) Try to extend as many identities for binomial coefficients as you can to Fibonomial coefficients.

(c) Generalize to (a,b)-recurrent sequences with arbitrary a and b.

5. Permutations

This chapter is devoted to permutations. We first recall how they are defined.

5.1. Permutations and the symmetric group

Definition 5.1. First, let us stipulate, once and for all, how we define the composition of two maps: If X, Y and Z are three sets, and if $\alpha : X \to Y$ and $\beta : Y \to Z$ are two maps, then $\beta \circ \alpha$ denotes the map from X to Z which sends every $x \in X$ to $\beta(\alpha(x))$. This map $\beta \circ \alpha$ is called the *composition* of β and α (and is sometimes abbreviated as $\beta\alpha$). This is the classical notation for composition of maps, and the reason why I am so explicitly reminding you of it is that some people (e.g., Herstein in [Herstei75]) use a different convention that conflicts with it: They write maps "on the right" (i.e., they denote the image of an element $x \in X$ under the map $\alpha : X \to Y$ by x^α or $x\alpha$ instead of $\alpha(x)$), and they define composition "the other way round" (i.e., they write $\alpha \circ \beta$ for what we call $\beta \circ \alpha$). They have reasons for what they are doing, but I shall use the classical notation because most of the literature agrees with it.

Definition 5.2. Let us also recall what it means for two maps to be *inverse*.

Let X and Y be two sets. Two maps $f : X \to Y$ and $g : Y \to X$ are said to be *mutually inverse* if they satisfy $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. (In other words, two maps $f : X \to Y$ and $g : Y \to X$ are mutually inverse if and only if every $x \in X$ satisfies $g(f(x)) = x$ and every $y \in Y$ satisfies $f(g(y)) = y$.)

Let $f : X \to Y$ be a map. If there exists a map $g : Y \to X$ such that f and g are mutually inverse, then this map g is unique (this is easy to check) and is called the *inverse* of f and denoted by f^{-1}. In this case, the map f is said to be *invertible*. It is easy to see that if g is the inverse of f, then f is the inverse of g.

It is well-known that a map $f : X \to Y$ is invertible if and only if f is bijective (i.e., both injective and surjective). The words "invertible" and "bijective" are thus synonyms (at least when used for a map between two sets – in other situations, they can be rather different). Nevertheless, both of them are commonly used, often by the same authors (since they convey slightly different mental images).

A bijective map is also called a *bijection* or a *1-to-1 correspondence* (or a *one-to-one correspondence*). When there is a bijection from X to Y, one says that the elements of X are *in bijection with* (or *in one-to-one correspondence with*) the elements of Y. It is well-known that two sets X and Y have the same cardinality if and only if there exists a bijection from X to Y. (This is precisely Theorem 1.2.)

Definition 5.3. A *permutation* of a set X means a bijection from X to X. The permutations of a given set X can be composed (i.e., if α and β are two permutations of X, then so is $\alpha \circ \beta$) and have inverses (which, again, are permutations of X). More precisely:

- If α and β are two permutations of a given set X, then the composition $\alpha \circ \beta$ is again a permutation of X.
- Any three permutations α, β and γ of X satisfy $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$. (This holds, more generally, for arbitrary maps which can be composed.)
- The identity map $\mathrm{id} : X \to X$ (this is the map which sends every element $x \in X$ to itself) is a permutation of X; it is also called the *identity permutation*. Every permutation α of X satisfies $\mathrm{id} \circ \alpha = \alpha$ and $\alpha \circ \mathrm{id} = \alpha$. (Again, this can be generalized to arbitrary maps.)
- For every permutation α of X, the inverse map α^{-1} is well-defined and is again a permutation of X. We have $\alpha \circ \alpha^{-1} = \mathrm{id}$ and $\alpha^{-1} \circ \alpha = \mathrm{id}$.

In the lingo of algebraists, these four properties show that the set of all permutations of X is a group whose binary operation is composition, and whose neutral element is the identity permutation $\mathrm{id} : X \to X$. This group is known as the *symmetric group of the set* X. (We will define the notion of a group later, in Definition 6.116; thus you might not understand the preceding two sentences at this point. If you do not care about groups, you should just remember that the symmetric group of X is the set of all permutations of X.)

Remark 5.4. Some authors define a permutation of a finite set X to mean a list of all elements of X, each occurring exactly once. This is **not** the meaning that the word "permutation" has in these notes! It is a different notion which, for historical reasons, has been called "permutation" as well. On the Wikipedia page for "permutation", the two notions are called "active" and "passive", respectively: An "active" permutation of X means a bijection from X to X (that is, a permutation of X in our meaning of this word), whereas a "passive" permutation of X means a list of all elements of X, each occurring exactly once. For example, if $X = \{\text{"cat"}, \text{"dog"}, \text{"archaeopteryx"}\}$, then the map

$$\begin{aligned} \text{"cat"} &\mapsto \text{"archaeopteryx"}, \\ \text{"archaeopteryx"} &\mapsto \text{"dog"}, \\ \text{"dog"} &\mapsto \text{"cat"} \end{aligned}$$

is an "active" permutation of X, whereas the list ("dog", "cat", "archaeopteryx") is a "passive" permutation of X.

When X is the set $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, then it is possible to equate each "active" permutation of X with a "passive" permutation of X (namely, its one-line notation, defined below). More generally, this can be done when X comes with a fixed total order. In general, if X is a finite set, then the number of "active" permutations of X equals the number of "passive" permutations of X (and both numbers equal $|X|!$), but until you fix some ordering of the elements of X, there is no "natural" way to match the "passive" permutations with the "active" ones. (And when X is infinite, the notion of a "passive" permutation is not even well-defined.)

To reiterate: For us, the word "permutation" shall always mean an "active" permutation!

Recall that $\mathbb{N} = \{0, 1, 2, \ldots\}$.

Definition 5.5. Let $n \in \mathbb{N}$.

Let S_n be the symmetric group of the set $\{1, 2, \ldots, n\}$. This is the set of all permutations of the set $\{1, 2, \ldots, n\}$. It contains the identity permutation $\mathrm{id} \in S_n$ which sends every $i \in \{1, 2, \ldots, n\}$ to i.

A well-known fact states that for every $n \in \mathbb{N}$, the size of the symmetric group S_n is $|S_n| = n!$ (that is, there are exactly $n!$ permutations of $\{1, 2, \ldots, n\}$). (One proof of this fact – not the simplest – is given in the proof of Corollary 7.81 below.)

We will often write a permutation $\sigma \in S_n$ as the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ of its values. This is known as the *one-line notation* for permutations (because it is a single-rowed list, as opposed to e.g. the two-line notation which is a two-rowed table).[142] For instance, the permutation in S_3 which sends 1 to 2, 2 to 1 and 3 to 3 is written $(2, 1, 3)$ in one-line notation.

The exact relation between lists and permutations is given by the following simple fact:

[142] Combinatorialists often omit the parentheses and the commas (i.e., they just write $\sigma(1)\sigma(2)\cdots\sigma(n)$, hoping that noone will mistake this for a product), since there is unfortunately another notation for permutations (the *cycle notation*) which also writes them as lists (actually, lists of lists) but where the lists have a different meaning.

Proposition 5.6. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$.

(a) If $\sigma \in S_n$, then each element of $[n]$ appears exactly once in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$.

(b) If $(p_1, p_2, \ldots, p_n)$ is a list of elements of $[n]$ such that each element of $[n]$ appears exactly once in this list $(p_1, p_2, \ldots, p_n)$, then there exists a unique permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$.

(c) Let $k \in \{0, 1, \ldots, n\}$. If $(p_1, p_2, \ldots, p_k)$ is a list of some elements of $[n]$ such that $p_1, p_2, \ldots, p_k$ are distinct, then there exists a permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_k) = (\sigma(1), \sigma(2), \ldots, \sigma(k))$.

At this point, let us clarify what we mean by "distinct": Several objects $u_1, u_2, \ldots, u_k$ are said to be *distinct* if every $i \in \{1, 2, \ldots, k\}$ and $j \in \{1, 2, \ldots, k\}$ satisfying $i \neq j$ satisfy $u_i \neq u_j$. (Some people call this "pairwise distinct".) So, for example, the numbers $2, 1, 6$ are distinct, but the numbers $6, 1, 6$ are not (although 6 and 1 are distinct). Instead of saying that some objects $u_1, u_2, \ldots, u_k$ are distinct, we can also say that "the list $(u_1, u_2, \ldots, u_k)$ has no repetitions"[143].

Remark 5.7. The σ in Proposition 5.6 **(b)** is uniquely determined, but the σ in Proposition 5.6 **(c)** is not (in general). More precisely, in Proposition 5.6 **(c)**, there are $(n-k)!$ possible choices of σ that work. (This is easy to check.)

Proof of Proposition 5.6. Proposition 5.6 is a basic fact and its proof is simple. I am going to present the proof in great detail, but you are not missing much if you skip it for its obviousness (just make sure you know **why** it is obvious).

Recall that S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$. In other words, S_n is the set of all permutations of the set $[n]$ (since $\{1, 2, \ldots, n\} = [n]$).

(a) Let $\sigma \in S_n$. Let $i \in [n]$.

We have $\sigma \in S_n$. In other words, σ is a permutation of $[n]$ (since S_n is the set of all permutations of the set $[n]$). In other words, σ is a bijective map $[n] \to [n]$. Hence, σ is both surjective and injective.

Now, we make the following two observations:

- The number i appears in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ [144].
- The number i appears at most once in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ [145].

[143] A repetition just means an element which occurs more than once in the list. It does not matter whether the occurrences are at consecutive positions or not.

[144] *Proof.* The map σ is surjective. Hence, there exists some $j \in [n]$ such that $i = \sigma(j)$. In other words, the number i appears in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. Qed.

[145] *Proof.* Let us assume the contrary (for the sake of contradiction). Thus, i appears more than once in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. In other words, i appears at least twice in this list. In other words, there exist two distinct elements p and q of $[n]$ such that $\sigma(p) = i$ and $\sigma(q) = i$. Consider these p and q.

We have $p \neq q$ (since p and q are distinct), so that $\sigma(p) \neq \sigma(q)$ (since σ is injective). This contradicts $\sigma(p) = i = \sigma(q)$. This contradiction proves that our assumption was wrong, qed.

Combining these two observations, we conclude that the number i appears exactly once in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$.

Let us now forget that we fixed i. We thus have shown that if $i \in [n]$, then i appears exactly once in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. In other words, each element of $[n]$ appears exactly once in the list $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. This proves Proposition 5.6 **(a)**.

(b) Let $(p_1, p_2, \ldots, p_n)$ be a list of elements of $[n]$ such that each element of $[n]$ appears exactly once in this list $(p_1, p_2, \ldots, p_n)$.

We have $p_i \in [n]$ for every $i \in [n]$ (since $(p_1, p_2, \ldots, p_n)$ is a list of elements of $[n]$).

We define a map $\tau : [n] \to [n]$ by setting

$$(\tau(i) = p_i \qquad \text{for every } i \in [n]). \tag{292}$$

(This is well-defined, because we have $p_i \in [n]$ for every $i \in [n]$.) The map τ is injective[146] and surjective[147]. Hence, the map τ is bijective. In other words, τ is a permutation of $[n]$ (since τ is a map $[n] \to [n]$). In other words, $\tau \in S_n$ (since S_n is the set of all permutations of the set $[n]$). Clearly, $(\tau(1), \tau(2), \ldots, \tau(n)) = (p_1, p_2, \ldots, p_n)$ (because of (292)), so that $(p_1, p_2, \ldots, p_n) = (\tau(1), \tau(2), \ldots, \tau(n))$.

Hence, there exists a permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ (namely, $\sigma = \tau$). Moreover, there exists **at most one** such permutation[148]. Combining the claims of the previous two

---

[146] *Proof.* Let u and v be two elements of $[n]$ such that $\tau(u) = \tau(v)$. We shall show that $u = v$.

Indeed, we assume the contrary (for the sake of contradiction). Thus, $u \neq v$.

The definition of $\tau(u)$ shows that $\tau(u) = p_u$. But we also have $\tau(u) = \tau(v) = p_v$ (by the definition of $\tau(v)$). Now, the element $\tau(u)$ of $[n]$ appears (at least) twice in the list $(p_1, p_2, \ldots, p_n)$: once at the u-th position (since $\tau(u) = p_u$), and again at the v-th position (since $\tau(u) = p_v$). (And these are two distinct positions, because $u \neq v$.)

But let us recall that each element of $[n]$ appears exactly once in this list $(p_1, p_2, \ldots, p_n)$. Hence, no element of $[n]$ appears more than once in the list $(p_1, p_2, \ldots, p_n)$. In particular, $\tau(u)$ cannot appear more than once in this list $(p_1, p_2, \ldots, p_n)$. This contradicts the fact that $\tau(u)$ appears twice in the list $(p_1, p_2, \ldots, p_n)$.

This contradiction shows that our assumption was wrong. Hence, $u = v$ is proven.

Now, let us forget that we fixed u and v. We thus have proven that if u and v are two elements of $[n]$ such that $\tau(u) = \tau(v)$, then $u = v$. In other words, the map τ is injective. Qed.

[147] *Proof.* Let $u \in [n]$. Each element of $[n]$ appears exactly once in the list $(p_1, p_2, \ldots, p_n)$. Applying this to the element u of $[n]$, we conclude that u appears exactly once in the list $(p_1, p_2, \ldots, p_n)$. In other words, there exists exactly one $i \in [n]$ such that $u = p_i$. Consider this i. The definition of τ yields $\tau(i) = p_i$. Compared with $u = p_i$, this yields $\tau(i) = u$.

Hence, there exists a $j \in [n]$ such that $\tau(j) = u$ (namely, $j = i$).

Let us now forget that we fixed u. We thus have proven that for every $u \in [n]$, there exists a $j \in [n]$ such that $\tau(j) = u$. In other words, the map τ is surjective. Qed.

[148] *Proof.* Let σ_1 and σ_2 be two permutations $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. Thus, σ_1 is a permutation in S_n such that $(p_1, p_2, \ldots, p_n) = (\sigma_1(1), \sigma_1(2), \ldots, \sigma_1(n))$, and σ_2 is a permutation in S_n such that $(p_1, p_2, \ldots, p_n) = (\sigma_2(1), \sigma_2(2), \ldots, \sigma_2(n))$.

We have $(\sigma_1(1), \sigma_1(2), \ldots, \sigma_1(n)) = (p_1, p_2, \ldots, p_n) = (\sigma_2(1), \sigma_2(2), \ldots, \sigma_2(n))$. In other words, every $i \in [n]$ satisfies $\sigma_1(i) = \sigma_2(i)$. In other words, $\sigma_1 = \sigma_2$.

sentences, we conclude that there exists a unique permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. This proves Proposition 5.6 **(b)**.

(c) Let $(p_1, p_2, \ldots, p_k)$ be a list of some elements of $[n]$ such that $p_1, p_2, \ldots, p_k$ are distinct. Thus, the list $(p_1, p_2, \ldots, p_k)$ contains k of the n elements of $[n]$ (because $p_1, p_2, \ldots, p_k$ are distinct). Let $q_1, q_2, \ldots, q_{n-k}$ be the remaining $n-k$ elements of $[n]$ (listed in any arbitrary order, with no repetition). Then, $(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_{n-k})$ is a list of all n elements of $[n]$, with no repetitions[149]. In other words, each element of $[n]$ appears exactly once in this list $(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_{n-k})$ (and each entry in this list is an element of $[n]$). Hence, we can apply Proposition 5.6 **(b)** to $(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_{n-k})$ instead of $(p_1, p_2, \ldots, p_n)$. As a consequence, we conclude that there exists a unique permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_{n-k}) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. Let τ be this σ.

Thus, $\tau \in S_n$ is a permutation such that

$$(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_{n-k}) = (\tau(1), \tau(2), \ldots, \tau(n)).$$

Now,

$$\begin{aligned} & (p_1, p_2, \ldots, p_k) \\ & = \left(\text{the list of the first } k \text{ entries of the list } \underbrace{(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_{n-k})}_{=(\tau(1), \tau(2), \ldots, \tau(n))} \right) \\ & = (\text{the list of the first } k \text{ entries of the list } (\tau(1), \tau(2), \ldots, \tau(n))) \\ & = (\tau(1), \tau(2), \ldots, \tau(k)). \end{aligned}$$

Hence, there exists a permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_k) = (\sigma(1), \sigma(2), \ldots, \sigma(k))$ (namely, $\sigma = \tau$). This proves Proposition 5.6 **(c)**. □

5.2. Inversions, lengths and the permutations $s_i \in S_n$

Let us now forget that we fixed σ_1 and σ_2. We thus have shown that if σ_1 and σ_2 are two permutations $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$, then $\sigma_1 = \sigma_2$. In other words, any two permutations $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ must be equal to each other. In other words, there exists **at most one** permutation $\sigma \in S_n$ such that $(p_1, p_2, \ldots, p_n) = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. Qed.

[149] It has no repetitions because:

- there are no repetitions among $p_1, p_2, \ldots, p_k$;
- there are no repetitions among $q_1, q_2, \ldots, q_{n-k}$;
- the two lists $(p_1, p_2, \ldots, p_k)$ and $(q_1, q_2, \ldots, q_{n-k})$ have no elements in common (because we defined $q_1, q_2, \ldots, q_{n-k}$ to be the "remaining" $n-k$ elements of $[n]$, where "remaining" means "not contained in the list $(p_1, p_2, \ldots, p_k)$").

Definition 5.8. Let $n \in \mathbb{N}$. For each $i \in \{1, 2, \ldots, n-1\}$, let s_i be the permutation in S_n that switches i with $i+1$ but leaves all other numbers unchanged. Formally speaking, s_i is the permutation in S_n given by

$$\left(s_i(k) = \begin{cases} i+1, & \text{if } k = i; \\ i, & \text{if } k = i+1; \\ k, & \text{if } k \notin \{i, i+1\} \end{cases} \qquad \text{for all } k \in \{1, 2, \ldots, n\} \right).$$

Thus, in one-line notation

$$s_i = (1, 2, \ldots, i-1, i+1, i, i+2, \ldots, n).$$

Notice that $s_i^2 = \text{id}$ for every $i \in \{1, 2, \ldots, n-1\}$. (Here, we are using the notation α^2 for $\alpha \circ \alpha$, where α is a permutation in S_n.)

Exercise 5.1. Let $n \in \mathbb{N}$.

(a) Show that $s_i \circ s_{i+1} \circ s_i = s_{i+1} \circ s_i \circ s_{i+1}$ for all $i \in \{1, 2, \ldots, n-2\}$.

(b) Show that every permutation $\sigma \in S_n$ can be written as a composition of several permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). For example, if $n = 3$, then the permutation[150] $(3, 1, 2)$ in S_3 can be written as the composition $s_2 \circ s_1$, while the permutation $(3, 2, 1)$ in S_3 can be written as the composition $s_1 \circ s_2 \circ s_1$ or also as the composition $s_2 \circ s_1 \circ s_2$.

[**Hint:** If you do not immediately see why this works, consider reading further.]

(c) Let w_0 denote the permutation in S_n which sends each $k \in \{1, 2, \ldots, n\}$ to $n+1-k$. (In one-line notation, this w_0 is written as $(n, n-1, \ldots, 1)$.) Find an **explicit** way to write w_0 as a composition of several permutations of the form s_i (with $i \in \{1, 2, \ldots, n-1\}$).

Remark 5.9. Symmetric groups appear in almost all parts of mathematics; unsurprisingly, there is no universally accepted notation for them. We are using the notation S_n for the n-th symmetric group; other common notations for it are $\mathfrak{S}_n$, Σ_n and $\operatorname{Sym}(n)$. The permutations that we call $s_1, s_2, \ldots, s_{n-1}$ are often called $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$. As already mentioned in Definition 5.1, some people write the composition of maps "backwards", which causes their $\sigma \circ \tau$ to be our $\tau \circ \sigma$, etc.. (Sadly, most authors are so sure that their notation is standard that they never bother to define it.)

In the language of group theory, the statement of Exercise 5.1 **(b)** says (or, more precisely, yields) that the permutations $s_1, s_2, \ldots, s_{n-1}$ generate the group S_n.

[150] Recall that we are writing permutations in one-line notation. Thus, "the permutation $(3, 1, 2)$ in S_3" means the permutation $\sigma \in S_3$ satisfying $(\sigma(1), \sigma(2), \sigma(3)) = (3, 1, 2)$.

Definition 5.10. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ be a permutation.

(a) An *inversion* of σ means a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. For instance, the inversions of the permutation $(3, 1, 2)$ (again, shown here in one-line notation) in S_3 are $(1, 2)$ and $(1, 3)$ (because $3 > 1$ and $3 > 2$), while the only inversion of the permutation $(1, 3, 2)$ in S_3 is $(2, 3)$ (since $3 > 2$).

(b) The *length* of σ means the number of inversions of σ. This length is denoted by $\ell(\sigma)$; it is a nonnegative integer.

If $n \in \mathbb{N}$, then any $\sigma \in S_n$ satisfies $0 \leq \ell(\sigma) \leq \dbinom{n}{2}$ (since the number of inversions of σ is clearly no larger than the total number of pairs (i, j) of integers satisfying $1 \leq i < j \leq n$; but the latter number is $\dbinom{n}{2}$). The only permutation in S_n having length 0 is the identity permutation $\mathrm{id} = (1, 2, \ldots, n) \in S_n$ [151].

Exercise 5.2. Let $n \in \mathbb{N}$.

(a) Show that every permutation $\sigma \in S_n$ and every $k \in \{1, 2, \ldots, n-1\}$ satisfy

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases} \tag{293}$$

and

$$\ell(s_k \circ \sigma) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1) \end{cases}. \tag{294}$$

(b) Show that any two permutations σ and τ in S_n satisfy $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2$.

(c) Show that any two permutations σ and τ in S_n satisfy $\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau)$.

(d) If $\sigma \in S_n$ is a permutation satisfying $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$, then show that $\sigma = \mathrm{id}$.

(e) Let $\sigma \in S_n$. Show that σ can be written as a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$).

(f) Let $\sigma \in S_n$. Then, show that $\ell(\sigma) = \ell(\sigma^{-1})$.

(g) Let $\sigma \in S_n$. Show that $\ell(\sigma)$ is the smallest $N \in \mathbb{N}$ such that σ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$).

Example 5.11. Let us justify Exercise 5.2 **(a)** on an example. The solution to Exercise 5.2 **(a)** given below is essentially a (tiresome) formalization of the ideas seen in this example.

Let $n = 5$, $k = 3$ and $\sigma = (4, 2, 1, 5, 3)$ (written in one-line notation). Then, $\sigma \circ s_k = (4, 2, 5, 1, 3)$; this is the permutation obtained by switching the k-th and

[151] The fact that the identity permutation $\mathrm{id} \in S_n$ has length $\ell(\mathrm{id}) = 0$ is trivial. The fact that it is the only one such permutation is easy (it essentially follows from Exercise 5.2 **(d)**).

the $(k+1)$-th entry of σ (where the word "entry" refers to the one-line notation). On the other hand, $s_k \circ \sigma = (3,2,1,5,4)$; this is the permutation obtained by switching the entry k with the entry $k+1$ of σ. Mind the difference between these two operations.

The inversions of $\sigma = (4,2,1,5,3)$ are $(1,2)$, $(1,3)$, $(1,5)$, $(2,3)$ and $(4,5)$. These are the pairs (i,j) of positions such that i is before j (that is, $i < j$) but the i-th entry of σ is larger than the j-th entry of σ (that is, $\sigma(i) > \sigma(j)$). In other words, these are the pairs of positions at which the entries of σ are out of order. On the other hand, the inversions of $s_k \circ \sigma = (3,2,1,5,4)$ are $(1,2)$, $(1,3)$, $(2,3)$ and $(4,5)$. These are precisely the inversions of σ except for $(1,5)$. This is no surprise: In fact, $s_k \circ \sigma$ is obtained from σ by switching the entry k with the entry $k+1$, and this operation clearly preserves all inversions other than the one that is directly being turned around (i.e., the inversion (i,j) where $\{\sigma(i), \sigma(j)\} = \{k, k+1\}$; in our case, this is the inversion $(1,5)$). In general, when $\sigma^{-1}(k) > \sigma^{-1}(k+1)$ (that is, when k appears further left than $k+1$ in the one-line notation of σ), the inversions of $s_k \circ \sigma$ are the inversions of σ except for $\left(\sigma^{-1}(k+1), \sigma^{-1}(k)\right)$. Therefore, in this case, the number of inversions of $s_k \circ \sigma$ equals the number of inversions of σ plus 1. That is, in this case, $\ell(s_k \circ \sigma) = \ell(\sigma) + 1$. When $\sigma^{-1}(k) < \sigma^{-1}(k+1)$, a similar argument shows $\ell(s_k \circ \sigma) = \ell(\sigma) - 1$. This explains why (294) holds (although formalizing this argument will be tedious).

The inversions of $\sigma \circ s_k = (4,2,5,1,3)$ are $(1,2)$, $(1,4)$, $(1,5)$, $(2,4)$, $(3,4)$ and $(3,5)$. Unlike the inversions of $s_k \circ \sigma$, these are not directly related to the inversions of σ, so the argument in the previous paragraph does not prove (293). However, instead of considering inversions of σ, one can consider inversions of σ^{-1}. These are even more intuitive: They are the pairs of integers (i,j) with $1 \leq i < j \leq n$ such that i appears further right than j in the one-line notation of σ. For instance, the inversions of σ^{-1} are $(1,2)$, $(1,4)$, $(2,4)$, $(3,4)$ and $(3,5)$, whereas the inversions of $(\sigma \circ s_k)^{-1}$ are all of these and also $(1,5)$. But there is no need to repeat our proof of (294); it is easier to deduce (293) from (294) by applying (294) to σ^{-1} instead of σ and appealing to Exercise 5.2 **(f)**. (Again, see the solution below for the details.)

Notice that Exercise 5.2 **(e)** immediately yields Exercise 5.1 **(b)**.

Remark 5.12. When $n = 0$ or $n = 1$, we have $\{1, 2, \ldots, n-1\} = \varnothing$. Hence, Exercise 5.1 **(e)** looks strange in the case when $n = 0$ or $n = 1$, because in this case, there are no permutations of the form s_k to begin with. Nevertheless, it is correct. Indeed, when $n = 0$ or $n = 1$, there is only one permutation $\sigma \in S_n$, namely the identity permutation id, and it has length $\ell(\sigma) = \ell(\mathrm{id}) = 0$. Thus, in this case, Exercise 5.1 **(e)** claims that id can be written as a composition of 0 permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). This is true: Even from an empty set we can always pick 0 elements; and the composition of 0 permutations will be id.

Remark 5.13. The word "length" for $\ell(\sigma)$ can be confusing: It does not refer to the length of the n-tuple $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ (which is n). The reason why it is called "length" is Exercise 5.2 **(g)**: it says that $\ell(\sigma)$ is the smallest number of permutations of the form s_k which can be multiplied to give σ; thus, it is the smallest possible length of an expression of σ as a product of s_k's.

The use of the word "length", unfortunately, is not standard across literature. Some authors call "Coxeter length" what we call "length", and use the word "length" itself for a different notion.

Exercise 5.3. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. In Exercise 5.1 **(b)**, we have seen that σ can be written as a composition of several permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). Usually there will be several ways to do so (for instance, $\operatorname{id} = s_1 \circ s_1 = s_2 \circ s_2 = \cdots = s_{n-1} \circ s_{n-1}$). Show that, whichever of these ways we take, the number of permutations composed will be congruent to $\ell(\sigma)$ modulo 2.

5.3. The sign of a permutation

Definition 5.14. Let $n \in \mathbb{N}$.

(a) We define the *sign* of a permutation $\sigma \in S_n$ as the integer $(-1)^{\ell(\sigma)}$. We denote this sign by $(-1)^\sigma$ or $\operatorname{sign} \sigma$ or $\operatorname{sgn} \sigma$.

(b) We say that a permutation σ is *even* if its sign is 1 (that is, if $\ell(\sigma)$ is even), and *odd* if its sign is -1 (that is, if $\ell(\sigma)$ is odd).

Signs of permutations have the following properties:

Proposition 5.15. Let $n \in \mathbb{N}$.

(a) The sign of the identity permutation $\operatorname{id} \in S_n$ is $(-1)^{\operatorname{id}} = 1$. In other words, $\operatorname{id} \in S_n$ is even.

(b) For every $k \in \{1, 2, \ldots, n-1\}$, the sign of the permutation $s_k \in S_n$ is $(-1)^{s_k} = -1$.

(c) If σ and τ are two permutations in S_n, then $(-1)^{\sigma \circ \tau} = (-1)^\sigma \cdot (-1)^\tau$.

(d) If $\sigma \in S_n$, then $(-1)^{\sigma^{-1}} = (-1)^\sigma$. (Here and in the following, the expression "$(-1)^{\sigma^{-1}}$" should be read as "$(-1)^{(\sigma^{-1})}$", not as "$\left((-1)^\sigma\right)^{-1}$"; this is similar to Convention 2.58 (although σ is not a number).)

Proof of Proposition 5.15. **(a)** The identity permutation id satisfies $\ell(\operatorname{id}) = 0$ [152]. Now, the definition of $(-1)^{\operatorname{id}}$ yields $(-1)^{\operatorname{id}} = (-1)^{\ell(\operatorname{id})} = 1$ (since $\ell(\operatorname{id}) = 0$). In other words, $\operatorname{id} \in S_n$ is even. This proves Proposition 5.15 **(a)**.

[152] *Proof.* An inversion of id is the same as a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\operatorname{id}(i) > \operatorname{id}(j)$ (by the definition of "inversion"). Thus, if (i, j) is an inversion of id, then $1 \leq i < j \leq n$ and $\operatorname{id}(i) > \operatorname{id}(j)$; but this leads to a contradiction (since $\operatorname{id}(i) > \operatorname{id}(j)$ contradicts $\operatorname{id}(i) = i < j = \operatorname{id}(j)$). Hence, we obtain a contradiction for each inversion of id. Thus, there are no inversions of id. But $\ell(\operatorname{id})$ is defined as the number of inversions of id. Hence, $\ell(\operatorname{id}) = (\text{the number of inversions of } \operatorname{id}) = 0$ (since there are no inversions of id).

(b) Let $k \in \{1, 2, \ldots, n-1\}$. Applying (293) to $\sigma = \operatorname{id}$, we obtain

$$\begin{aligned}\ell(\operatorname{id} \circ s_k) &= \begin{cases} \ell(\operatorname{id}) + 1, & \text{if } \operatorname{id}(k) < \operatorname{id}(k+1); \\ \ell(\operatorname{id}) - 1, & \text{if } \operatorname{id}(k) > \operatorname{id}(k+1) \end{cases} \\ &= \underbrace{\ell(\operatorname{id})}_{=0} + 1 \qquad (\text{since } \operatorname{id}(k) = k < k+1 = \operatorname{id}(k+1)) \\ &= 1.\end{aligned}$$

This rewrites as $\ell(s_k) = 1$ (since $\operatorname{id} \circ s_k = s_k$). Now, the definition of $(-1)^{s_k}$ yields $(-1)^{s_k} = (-1)^{\ell(s_k)} = -1$ (since $\ell(s_k) = 1$). This proves Proposition 5.15 **(b)**.

(c) Let $\sigma \in S_n$ and $\tau \in S_n$. Exercise 5.2 **(b)** yields $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2$, so that $(-1)^{\ell(\sigma\circ\tau)} = (-1)^{\ell(\sigma)+\ell(\tau)} = (-1)^{\ell(\sigma)} \cdot (-1)^{\ell(\tau)}$. But the definition of the sign of a permutation yields $(-1)^{\sigma\circ\tau} = (-1)^{\ell(\sigma\circ\tau)}$ and $(-1)^{\sigma} = (-1)^{\ell(\sigma)}$ and $(-1)^{\tau} = (-1)^{\ell(\tau)}$. Hence, $(-1)^{\sigma\circ\tau} = (-1)^{\ell(\sigma\circ\tau)} = \underbrace{(-1)^{\ell(\sigma)}}_{=(-1)^{\sigma}} \cdot \underbrace{(-1)^{\ell(\tau)}}_{=(-1)^{\tau}} = (-1)^{\sigma} \cdot (-1)^{\tau}$. This proves Proposition 5.15 **(c)**.

(d) Let $\sigma \in S_n$. The definition of $(-1)^{\sigma^{-1}}$ yields $(-1)^{\sigma^{-1}} = (-1)^{\ell(\sigma^{-1})}$. But Exercise 5.2 **(f)** says that $\ell(\sigma) = \ell(\sigma^{-1})$. The definition of $(-1)^{\sigma}$ yields $(-1)^{\sigma} = (-1)^{\ell(\sigma)} = (-1)^{\ell(\sigma^{-1})}$ (since $\ell(\sigma) = \ell(\sigma^{-1})$). Compared with $(-1)^{\sigma^{-1}} = (-1)^{\ell(\sigma^{-1})}$, this yields $(-1)^{\sigma^{-1}} = (-1)^{\sigma}$. This proves Proposition 5.15 **(d)**. $\square$

If you are familiar with some basic concepts of abstract algebra, then you will immediately notice that parts **(a)** and **(c)** of Proposition 5.15 can be summarized as the statement that "sign is a group homomorphism from the group S_n to the multiplicative group $\{1, -1\}$". In this statement, "sign" means the map from S_n to $\{1, -1\}$ which sends every permutation σ to its sign $(-1)^{\sigma} = (-1)^{\ell(\sigma)}$, and the "multiplicative group $\{1, -1\}$" means the group $\{1, -1\}$ whose binary operation is multiplication.

We have defined the sign of a permutation $\sigma \in S_n$. More generally, it is possible to define the sign of a permutation of an arbitrary finite set X, even though the length of such a permutation is not defined![153]

[153]How does it work? If X is a finite set, then we can always find a bijection $\phi : X \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. (Constructing such a bijection is tantamount to writing down a list of all elements of X, with no duplicates.) Given such a bijection ϕ, we can define the sign of any permutation σ of X as follows:

$$(-1)^{\sigma} = (-1)^{\phi\circ\sigma\circ\phi^{-1}}. \qquad (295)$$

Here, the right hand side is well-defined because $\phi \circ \sigma \circ \phi^{-1}$ is a permutation of $\{1, 2, \ldots, n\}$. What is not immediately obvious is that this sign is independent on the choice of ϕ, and that it is a group homomorphism to $\{1, -1\}$ (that is, we have $(-1)^{\operatorname{id}} = 1$ and $(-1)^{\sigma\circ\tau} = (-1)^{\sigma} \cdot (-1)^{\tau}$). We will prove these facts further below (in Exercise 5.12).

Exercise 5.4. Let $n \geq 2$. Show that the number of even permutations in S_n is $n!/2$, and the number of odd permutations in S_n is also $n!/2$.

The sign of a permutation is used in the combinatorial definition of the determinant. Let us briefly show this definition now; we shall return to it later (in Chapter 6) to study it in much more detail.

Definition 5.16. Let $n \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix (say, with complex entries, although this does not matter much – it suffices that the entries can be added and multiplied and the axioms of associativity, distributivity, commutativity, unity etc. hold). The *determinant* $\det A$ of A is defined as

$$\sum_{\sigma \in S_n} (-1)^{\sigma} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}. \tag{296}$$

Let me try to describe the sum (296) in slightly more visual terms: The sum (296) has $n!$ addends, each of which has the form "$(-1)^{\sigma}$ times a product". The product has n factors, which are entries of A, and are chosen in such a way that there is exactly one entry taken from each row and exactly one from each column. Which precise entries are taken depends on σ: namely, for each i, we take the $\sigma(i)$-th entry from the i-th row.

Convince yourself that the classical formulas

$$\det \begin{pmatrix} a \end{pmatrix} = a;$$
$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc;$$
$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - ahf - bdi - ceg$$

are particular cases of (296). Whenever $n \geq 2$, the sum in (296) contains precisely $n!/2$ plus signs and $n!/2$ minus signs (because of Exercise 5.4).

Definition 5.16 is merely one of several equivalent definitions of the determinant. You will probably see two of them in an average linear algebra class. Each of them has its own advantages and drawbacks. Definition 5.16 is the most direct, assuming that one knows about the sign of a permutation.

5.4. Infinite permutations

(This section is optional; it explores some technical material which is useful in combinatorics, but is not necessary for what follows. I advise the reader to skip it at the first read.)

We have introduced the notion of a permutation of an arbitrary set; but so far, we have only studied permutations of finite sets. In this section (which is tangential to our project; probably nothing from this section will be used ever after), let me discuss permutations of the infinite set $\{1,2,3,\ldots\}$. (A lot of what I say below can be easily adapted to the sets $\mathbb{N}$ and $\mathbb{Z}$ as well.)

We recall that a permutation of a set X means a bijection from X to X.

Let S_∞ be the symmetric group of the set $\{1,2,3,\ldots\}$. This is the set of all permutations of $\{1,2,3,\ldots\}$. It contains the identity permutation $\operatorname{id} \in S_\infty$ which sends every $i \in \{1,2,3,\ldots\}$ to i. The set S_∞ is uncountable[154].

We shall try to study S_∞ similarly to how we studied S_n for $n \in \mathbb{N}$. However, we soon will notice that the analogy between S_∞ and S_n will break down.[155] To amend this, we shall define a subset $S_{(\infty)}$ of S_∞ (mind the parentheses around the "∞") which is smaller and more wieldy, and indeed shares many of the properties of the finite symmetric group S_n.

We define $S_{(\infty)}$ as follows:

$$S_{(\infty)} = \{\sigma \in S_\infty \mid \sigma(i) = i \text{ for all but finitely many } i \in \{1,2,3,\ldots\}\}. \tag{297}$$

Let us first explain what "all but finitely many $i \in \{1,2,3,\ldots\}$" means:

Definition 5.17. Let I be a set. Let $\mathcal{A}(i)$ be a statement for every $i \in I$. Then, we say that "$\mathcal{A}(i)$ for all but finitely many $i \in I$" if and only if there exists some finite subset J of I such that every $i \in I \setminus J$ satisfies $\mathcal{A}(i)$. [156]

Thus, for a permutation $\sigma \in S_\infty$, we have the following equivalence of statements:

$$\begin{aligned}
&(\sigma(i) = i \text{ for all but finitely many } i \in \{1,2,3,\ldots\}) \\
\iff &(\text{there exists some finite subset } J \text{ of } \{1,2,3,\ldots\} \text{ such that} \\
&\quad \text{every } i \in \{1,2,3,\ldots\} \setminus J \text{ satisfies } \sigma(i) = i) \\
\iff &(\text{there exists some finite subset } J \text{ of } \{1,2,3,\ldots\} \text{ such that} \\
&\quad \text{the only } i \in \{1,2,3,\ldots\} \text{ that satisfy } \sigma(i) \neq i \text{ are elements of } J) \\
\iff &(\text{the set of all } i \in \{1,2,3,\ldots\} \text{ that satisfy } \sigma(i) \neq i \text{ is} \\
&\quad \text{contained in some finite subset } J \text{ of } \{1,2,3,\ldots\}) \\
\iff &(\text{there are only finitely many } i \in \{1,2,3,\ldots\} \text{ that satisfy } \sigma(i) \neq i).
\end{aligned}$$

[154] More generally, while a finite set of size n has $n!$ permutations, an infinite set S has uncountably many permutations (even if S is countable).

[155] The uncountability of S_∞ is the first hint that S_∞ is "too large" a set to be a good analogue of the finite set S_n.

[156] Thus, the statement "$\mathcal{A}(i)$ for all but finitely many $i \in I$" can be restated as "$\mathcal{A}(i)$ holds for all $i \in I$, apart from finitely many exceptions" or as "there are only finitely many $i \in I$ which do not satisfy $\mathcal{A}(i)$". I prefer the first wording, because it makes the most sense in constructive logic.

Caution: Do not confuse the words "all but finitely many $i \in I$" in this definition with the words "infinitely many $i \in I$". For instance, it is true that n is even for infinitely many $n \in \mathbb{Z}$, but it is not true that n is even for all but finitely many $n \in \mathbb{Z}$. Conversely, it is true that $n > 1$ for all but finitely many $n \in \{1,2\}$ (because the only $n \in \{1,2\}$ which does not satisfy $n > 1$ is 1), but it is not true that $n > 1$ for infinitely many $n \in \{1,2\}$ (because there are no infinitely many $n \in \{1,2\}$ to begin with).

You will encounter the "all but finitely many" formulation often in abstract algebra. (Some people abbreviate it as "almost all", but this abbreviation means other things as well.)

Hence, (297) rewrites as follows:

$$S_{(\infty)} = \{\sigma \in S_\infty \mid \text{there are only finitely many } i \in \{1,2,3,\ldots\} \text{ that satisfy } \sigma(i) \neq i\}.$$

Example 5.18. Here is an example of a permutation which is in S_∞ but not in $S_{(\infty)}$: Let τ be the permutation of $\{1,2,3,\ldots\}$ given by

$$(\tau(1), \tau(2), \tau(3), \tau(4), \tau(5), \tau(6), \ldots) = (2,1,4,3,6,5,\ldots).$$

(It adds 1 to every odd positive integer, and subtracts 1 from every even positive integer.) Then, $\tau \in S_\infty$ but $\tau \notin S_{(\infty)}$.

On the other hand, let us show some examples of permutations in $S_{(\infty)}$. For each $i \in \{1,2,3,\ldots\}$, let s_i be the permutation in S_∞ that switches i with $i+1$ but leaves all other numbers unchanged. (This is similar to the permutation s_i in S_n that was defined earlier. We have taken the liberty to re-use the name s_i, hoping that no confusion will arise.)

Again, we have $s_i^2 = \mathrm{id}$ for every $i \in \{1,2,3,\ldots\}$ (where α^2 means $\alpha \circ \alpha$ for any $\alpha \in S_\infty$).

Proposition 5.19. We have $s_k \in S_{(\infty)}$ for every $k \in \{1,2,3,\ldots\}$.

Proof of Proposition 5.19. Let $k \in \{1,2,3,\ldots\}$. The permutation s_k has been defined as the permutation in S_∞ that switches k with $k+1$ but leaves all other numbers unchanged. In other words, it satisfies $s_k(k) = k+1$, $s_k(k+1) = k$ and

$$s_k(i) = i \qquad \text{for every } i \in \{1,2,3,\ldots\} \text{ such that } i \notin \{k, k+1\}. \qquad (298)$$

Now, every $i \in \{1,2,3,\ldots\} \setminus \{k, k+1\}$ satisfies $s_k(i) = i$ [157]. Hence, there exists some finite subset J of $\{1,2,3,\ldots\}$ such that every $i \in \{1,2,3,\ldots\} \setminus J$ satisfies $s_k(i) = i$ (namely, $J = \{k, k+1\}$). In other words, $s_k(i) = i$ for all but finitely many $i \in \{1,2,3,\ldots\}$.

Thus, s_k is an element of S_∞ satisfying $s_k(i) = i$ for all but finitely many $i \in \{1,2,3,\ldots\}$. Hence,

$$s_k \in \{\sigma \in S_\infty \mid \sigma(i) = i \text{ for all but finitely many } i \in \{1,2,3,\ldots\}\} = S_{(\infty)}.$$

This proves Proposition 5.19. □

Permutations can be composed and inverted, leading to new permutations. Let us first see that the same is true for elements of $S_{(\infty)}$:

[157] *Proof.* Let $i \in \{1,2,3,\ldots\} \setminus \{k, k+1\}$. Thus, $i \in \{1,2,3,\ldots\}$ and $i \notin \{k, k+1\}$. Hence, (298) shows that $s_k(i) = i$, qed.

Proposition 5.20. **(a)** The identity permutation $\operatorname{id} \in S_\infty$ of $\{1,2,3,\ldots\}$ satisfies $\operatorname{id} \in S_{(\infty)}$.

(b) For every $\sigma \in S_{(\infty)}$ and $\tau \in S_{(\infty)}$, we have $\sigma \circ \tau \in S_{(\infty)}$.

(c) For every $\sigma \in S_{(\infty)}$, we have $\sigma^{-1} \in S_{(\infty)}$.

Proof of Proposition 5.20. We have defined $S_{(\infty)}$ as the set of all $\sigma \in S_\infty$ such that $\sigma(i) = i$ for all but finitely many $i \in \{1,2,3,\ldots\}$. In other words, $S_{(\infty)}$ is the set of all $\sigma \in S_\infty$ such that there exists a finite subset K of $\{1,2,3,\ldots\}$ such that (every $i \in \{1,2,3,\ldots\} \setminus K$ satisfies $\sigma(i) = i$). As a consequence, we have the following two facts:

- If K is a finite subset of $\{1,2,3,\ldots\}$, and if $\gamma \in S_\infty$ is a permutation such that
$$\left(\text{every } i \in \{1,2,3,\ldots\} \setminus K \text{ satisfies } \gamma(i) = i\right), \tag{299}$$
then
$$\gamma \in S_{(\infty)}. \tag{300}$$

- If $\gamma \in S_{(\infty)}$, then
$$\left(\begin{array}{c} \text{there exists some finite subset } K \text{ of } \{1,2,3,\ldots\} \\ \text{such that every } i \in \{1,2,3,\ldots\} \setminus K \text{ satisfies } \gamma(i) = i \end{array}\right). \tag{301}$$

We can now step to the actual proof of Proposition 5.20.

(a) Every $i \in \{1,2,3,\ldots\} \setminus \varnothing$ satisfies $\operatorname{id}(i) = i$. Thus, (300) (applied to $K = \varnothing$ and $\gamma = \operatorname{id}$) yields $\operatorname{id} \in S_{(\infty)}$. This proves Proposition 5.20 **(a)**.

(b) Let $\sigma \in S_{(\infty)}$ and $\tau \in S_{(\infty)}$.

From (301) (applied to $\gamma = \sigma$), we conclude that there exists some finite subset K of $\{1,2,3,\ldots\}$ such that every $i \in \{1,2,3,\ldots\} \setminus K$ satisfies $\sigma(i) = i$. Let us denote this K by J_1. Thus, J_1 is a finite subset of $\{1,2,3,\ldots\}$, and
$$\text{every } i \in \{1,2,3,\ldots\} \setminus J_1 \text{ satisfies } \sigma(i) = i. \tag{302}$$

From (301) (applied to $\gamma = \tau$), we conclude that there exists some finite subset K of $\{1,2,3,\ldots\}$ such that every $i \in \{1,2,3,\ldots\} \setminus K$ satisfies $\tau(i) = i$. Let us denote this K by J_2. Thus, J_2 is a finite subset of $\{1,2,3,\ldots\}$, and
$$\text{every } i \in \{1,2,3,\ldots\} \setminus J_2 \text{ satisfies } \tau(i) = i. \tag{303}$$

The sets J_1 and J_2 are finite. Hence, their union $J_1 \cup J_2$ is finite. Moreover,
$$\text{every } i \in \{1,2,3,\ldots\} \setminus (J_1 \cup J_2) \text{ satisfies } (\sigma \circ \tau)(i) = i$$

[158]. Therefore, (300) (applied to $K = J_1 \cup J_2$ and $\gamma = \sigma \circ \tau$) yields $\sigma \circ \tau \in S_{(\infty)}$. This proves Proposition 5.20 **(b)**.

(c) Let $\sigma \in S_{(\infty)}$.

From (301) (applied to $\gamma = \sigma$), we conclude that there exists some finite subset K of $\{1, 2, 3, \ldots\}$ such that every $i \in \{1, 2, 3, \ldots\} \setminus K$ satisfies $\sigma(i) = i$. Consider this K. Thus, K is a finite subset of $\{1, 2, 3, \ldots\}$, and

$$\text{every } i \in \{1, 2, 3, \ldots\} \setminus K \text{ satisfies } \sigma(i) = i. \tag{304}$$

Now,

$$\text{every } i \in \{1, 2, 3, \ldots\} \setminus K \text{ satisfies } \sigma^{-1}(i) = i$$

[159]. Therefore, (300) (applied to $\gamma = \sigma^{-1}$) yields $\sigma^{-1} \in S_{(\infty)}$. This proves Proposition 5.20 **(c)**. □

In the language of group theorists, Proposition 5.20 show that $S_{(\infty)}$ is a subgroup of the group S_∞. The elements of $S_{(\infty)}$ are called the *finitary permutations of* $\{1, 2, 3, \ldots\}$, and $S_{(\infty)}$ is called the *finitary symmetric group of* $\{1, 2, 3, \ldots\}$.

We now have the following analogue of Exercise 5.1 (without its part **(c)**):

Exercise 5.5. **(a)** Show that $s_i \circ s_{i+1} \circ s_i = s_{i+1} \circ s_i \circ s_{i+1}$ for all $i \in \{1, 2, 3, \ldots\}$.

(b) Show that every permutation $\sigma \in S_{(\infty)}$ can be written as a composition of several permutations of the form s_k (with $k \in \{1, 2, 3, \ldots\}$).

Remark 5.21. In the language of group theory, the statement of Exercise 5.5 **(b)** says (or, more precisely, yields) that the permutations $s_1, s_2, s_3, \ldots$ generate the group $S_{(\infty)}$.

If $\sigma \in S_\infty$ is a permutation, then an *inversion* of σ means a pair (i, j) of integers satisfying $1 \leq i < j$ and $\sigma(i) > \sigma(j)$. This definition of an inversion is, of course, analogous to the definition of an inversion of a $\sigma \in S_n$; thus we could try to define the length of a $\sigma \in S_\infty$. However, here we run into troubles: A permutation $\sigma \in S_\infty$ might have infinitely many inversions!

It is here that we really need to restrict ourselves to $S_{(\infty)}$. This indeed helps:

Proposition 5.22. Let $\sigma \in S_{(\infty)}$. Then:

(a) There exists some $N \in \{1, 2, 3, \ldots\}$ such that every integer $i > N$ satisfies $\sigma(i) = i$.

(b) There are only finitely many inversions of σ.

[158] *Proof.* Let $i \in \{1, 2, 3, \ldots\} \setminus (J_1 \cup J_2)$. Thus, $i \in \{1, 2, 3, \ldots\}$ and $i \notin J_1 \cup J_2$.

We have $i \notin J_1 \cup J_2$ and thus $i \notin J_1$ (since $J_1 \subseteq J_1 \cup J_2$). Hence, $i \in \{1, 2, 3, \ldots\} \setminus J_1$. Similarly, $i \in \{1, 2, 3, \ldots\} \setminus J_2$. Thus, (303) yields $\tau(i) = i$. Hence, $(\sigma \circ \tau)(i) = \sigma\left(\underbrace{\tau(i)}_{=i}\right) = \sigma(i) = i$ (by (302)), qed.

[159] *Proof.* Let $i \in \{1, 2, 3, \ldots\} \setminus K$. Thus, $\sigma(i) = i$ (according to (304)), so that $\sigma^{-1}(i) = i$, qed.

Proof of Proposition 5.22. **(a)** We can apply (301) to $\gamma = \sigma$. As a consequence, we obtain that there exists some finite subset K of $\{1, 2, 3, \ldots\}$ such that

$$\text{every } i \in \{1, 2, 3, \ldots\} \setminus K \text{ satisfies } \sigma(i) = i. \tag{305}$$

Consider this K.

The set K is finite. Hence, the set $K \cup \{1\}$ is finite; this set is also nonempty (since it contains 1) and a subset of $\{1, 2, 3, \ldots\}$. Therefore, this set $K \cup \{1\}$ has a greatest element (since every finite nonempty subset of $\{1, 2, 3, \ldots\}$ has a greatest element). Let n be this greatest element. Clearly, $n \in K \cup \{1\} \subseteq \{1, 2, 3, \ldots\}$, so that $n > 0$.

Every $j \in K \cup \{1\}$ satisfies

$$j \leq n \tag{306}$$

(since n is the greatest element of $K \cup \{1\}$). Now, let i be an integer such that $i > n$. Then, $i > n > 0$, so that i is a positive integer. If we had $i \in K$, then we would have $i \in K \subseteq K \cup \{1\}$ and thus $i \leq n$ (by (306), applied to $j = i$), which would contradict $i > n$. Hence, we cannot have $i \in K$. We thus have $i \notin K$. Since $i \in \{1, 2, 3, \ldots\}$, this shows that $i \in \{1, 2, 3, \ldots\} \setminus K$. Thus, $\sigma(i) = i$ (by (305)).

Let us now forget that we fixed i. We thus have shown that every integer $i > n$ satisfies $\sigma(i) = i$. Hence, Proposition 5.22 **(a)** holds (we can take $N = n$).

(b) Proposition 5.22 **(a)** shows that there exists some $N \in \{1, 2, 3, \ldots\}$ such that

$$\text{every integer } i > N \text{ satisfies } \sigma(i) = i. \tag{307}$$

Consider such an N. We shall now show that

$$\text{every inversion of } \sigma \text{ is an element of } \{1, 2, \ldots, N\}^2.$$

In fact, let c be an inversion of σ. We will show that c is an element of $\{1, 2, \ldots, N\}^2$.

We know that c is an inversion of σ. In other words, c is a pair (i, j) of integers satisfying $1 \leq i < j$ and $\sigma(i) > \sigma(j)$ (by the definition of an "inversion of σ"). Consider this (i, j). We then have $i \leq N$ [160] and $j \leq N$ [161]. Consequently, $(i, j) \in \{1, 2, \ldots, N\}^2$. Hence, $c = (i, j) \in \{1, 2, \ldots, N\}^2$.

Now, let us forget that we fixed c. We thus have shown that if c is an inversion of σ, then c is an element of $\{1, 2, \ldots, N\}^2$. In other words, every inversion of σ is an element of $\{1, 2, \ldots, N\}^2$. Thus, there are only finitely many inversions of σ (since there are only finitely many elements of $\{1, 2, \ldots, N\}^2$). Proposition 5.22 **(b)** is thus proven. □

[160] *Proof.* Assume the contrary. Thus, $i > N$. Hence, (307) shows that $\sigma(i) = i$. Also, $i < j$, so that $j > i > N$. Hence, (307) (applied to j instead of i) shows that $\sigma(j) = j$. Thus, $\sigma(i) = i < j = \sigma(j)$. This contradicts $\sigma(i) > \sigma(j)$. This contradiction shows that our assumption was wrong, qed.

[161] *Proof.* Assume the contrary. Thus, $j > N$. Hence, (307) (applied to j instead of i) shows that $\sigma(j) = j$. Now, $\sigma(i) > \sigma(j) = j > N$. Therefore, (307) (applied to $\sigma(i)$ instead of i) yields $\sigma(\sigma(i)) = \sigma(i)$. But σ is a permutation, and thus an injective map. Hence, from $\sigma(\sigma(i)) = \sigma(i)$, we obtain $\sigma(i) = i$. Thus, $\sigma(i) = i < j = \sigma(j)$. This contradicts $\sigma(i) > \sigma(j)$. This contradiction shows that our assumption was wrong, qed.

Actually, Proposition 5.22 **(b)** has a converse: If a permutation $\sigma \in S_\infty$ has only finitely many inversions, then σ belongs to $S_{(\infty)}$. This is easy to prove; but we won't use this.

If $\sigma \in S_{(\infty)}$ is a permutation, then the *length* of σ means the number of inversions of σ. This is well-defined, because there are only finitely many inversions of σ (by Proposition 5.22 **(b)**). The length of σ is denoted by $\ell(\sigma)$; it is a nonnegative integer. The only permutation having length 0 is the identity permutation $\operatorname{id} \in S_\infty$.

We have the following analogue of Exercise 5.2:

Exercise 5.6. **(a)** Show that every permutation $\sigma \in S_{(\infty)}$ and every $k \in \{1, 2, 3, \ldots\}$ satisfy

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases} \tag{308}$$

and

$$\ell(s_k \circ \sigma) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1) \end{cases}. \tag{309}$$

(b) Show that any two permutations σ and τ in $S_{(\infty)}$ satisfy $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2$.

(c) Show that any two permutations σ and τ in $S_{(\infty)}$ satisfy $\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau)$.

(d) If $\sigma \in S_{(\infty)}$ is a permutation satisfying $\sigma(1) \leq \sigma(2) \leq \sigma(3) \leq \cdots$, then show that $\sigma = \operatorname{id}$.

(e) Let $\sigma \in S_{(\infty)}$. Show that σ can be written as a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, 3, \ldots\}$).

(f) Let $\sigma \in S_{(\infty)}$. Then, show that $\ell(\sigma) = \ell(\sigma^{-1})$.

(g) Let $\sigma \in S_{(\infty)}$. Show that $\ell(\sigma)$ is the smallest $N \in \mathbb{N}$ such that σ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, 3, \ldots\}$).

We also have an analogue of Exercise 5.3:

Exercise 5.7. Let $\sigma \in S_{(\infty)}$. In Exercise 5.5 **(b)**, we have seen that σ can be written as a composition of several permutations of the form s_k (with $k \in \{1, 2, 3, \ldots\}$). Usually there will be several ways to do so (for instance, $\operatorname{id} = s_1 \circ s_1 = s_2 \circ s_2 = s_3 \circ s_3 = \cdots$). Show that, whichever of these ways we take, the number of permutations composed will be congruent to $\ell(\sigma)$ modulo 2.

Having defined the length of a permutation $\sigma \in S_{(\infty)}$, we can now define the sign of such a permutation. Again, we mimic the definition of the sign of a $\sigma \in S_n$:

Definition 5.23. We define the *sign* of a permutation $\sigma \in S_{(\infty)}$ as the integer $(-1)^{\ell(\sigma)}$. We denote this sign by $(-1)^\sigma$ or $\operatorname{sign} \sigma$ or $\operatorname{sgn} \sigma$. We say that a permutation σ is *even* if its sign is 1 (that is, if $\ell(\sigma)$ is even), and *odd* if its sign is -1 (that is, if $\ell(\sigma)$ is odd).

Signs of permutations have the following properties:

Proposition 5.24. **(a)** The sign of the identity permutation $\operatorname{id} \in S_{(\infty)}$ is $(-1)^{\operatorname{id}} = 1$. In other words, $\operatorname{id} \in S_{(\infty)}$ is even.

(b) For every $k \in \{1, 2, 3, \ldots\}$, the sign of the permutation $s_k \in S_{(\infty)}$ is $(-1)^{s_k} = -1$.

(c) If σ and τ are two permutations in $S_{(\infty)}$, then $(-1)^{\sigma \circ \tau} = (-1)^{\sigma} \cdot (-1)^{\tau}$.

(d) If $\sigma \in S_{(\infty)}$, then $(-1)^{\sigma^{-1}} = (-1)^{\sigma}$.

The proof of Proposition 5.24 is analogous to the proof of Proposition 5.15.

Remark 5.25. We have defined the sign of a permutation $\sigma \in S_{(\infty)}$. No such notion exists for permutations $\sigma \in S_{\infty}$. In fact, one can show that if an element λ_σ of $\{1, -1\}$ is chosen for each $\sigma \in S_\infty$ in such a way that every two permutations $\sigma, \tau \in S_\infty$ satisfy $\lambda_{\sigma \circ \tau} = \lambda_\sigma \cdot \lambda_\tau$, then all of the λ_σ are 1. (Indeed, this follows from a result of Oystein Ore; see `http://mathoverflow.net/questions/54371` .)

Remark 5.26. For every $n \in \mathbb{N}$ and every $\sigma \in S_n$, we can define a permutation $\sigma_{(\infty)} \in S_{(\infty)}$ by setting

$$\sigma_{(\infty)}(i) = \begin{cases} \sigma(i), & \text{if } i \leq n; \\ i, & \text{if } i > n \end{cases} \qquad \text{for all } i \in \{1, 2, 3, \ldots\}.$$

Essentially, $\sigma_{(\infty)}$ is the permutation σ extended to the set of all positive integers in the laziest possible way: It just sends each $i > n$ to itself.

For every $n \in \mathbb{N}$, there is an injective map $\mathbf{i}_n : S_n \to S_{(\infty)}$ defined as follows:

$$\mathbf{i}_n(\sigma) = \sigma_{(\infty)} \qquad \text{for every } \sigma \in S_n.$$

This map $\mathbf{i}_n$ is an example of what algebraists call a *group homomorphism*: It satisfies

$$\begin{aligned} \mathbf{i}_n(\operatorname{id}) &= \operatorname{id}; \\ \mathbf{i}_n(\sigma \circ \tau) &= \mathbf{i}_n(\sigma) \circ \mathbf{i}_n(\tau) && \text{for all } \sigma, \tau \in S_n; \\ \mathbf{i}_n\left(\sigma^{-1}\right) &= (\mathbf{i}_n(\sigma))^{-1} && \text{for all } \sigma \in S_n. \end{aligned}$$

(This is all easy to check.) Thus, we can consider the image $\mathbf{i}_n(S_n)$ of S_n under this map as a "copy" of S_n which is "just as good as the original" (i.e., composition in this copy behaves in the same way as composition in the original). It is easy to characterize this copy inside $S_{(\infty)}$: Namely,

$$\mathbf{i}_n(S_n) = \left\{\sigma \in S_{(\infty)} \mid \sigma(i) = i \text{ for all } i > n\right\}.$$

Using Proposition 5.22 **(a)**, it is easy to check that $S_{(\infty)} = \bigcup_{n\in\mathbb{N}} \mathbf{i}_n(S_n) = \mathbf{i}_0(S_0) \cup \mathbf{i}_1(S_1) \cup \mathbf{i}_2(S_2) \cup \cdots$. Therefore, many properties of $S_{(\infty)}$ can be derived from analogous properties of S_n for finite n. For example, using this tactic, we could easily derive Exercise 5.6 from Exercise 5.2, and derive Exercise 5.7 from Exercise 5.3. (However, we can just as well solve Exercises 5.6 and 5.7 by copying the solutions of Exercises 5.2 and 5.3 and making the necessary changes; this is how I solve these exercises further below.)

5.5. More on lengths of permutations

Let us summarize some of what we have learnt about permutations. We have defined the length $\ell(\sigma)$ and the inversions of a permutation $\sigma \in S_n$, where n is a nonnegative integer. We recall the basic properties of these objects:

- For each $k \in \{1, 2, \ldots, n-1\}$, we defined s_k to be the permutation in S_n that switches k with $k+1$ but leaves all other numbers unchanged. These permutations satisfy $s_i^2 = \mathrm{id}$ for every $i \in \{1, 2, \ldots, n-1\}$ and

$$s_i \circ s_{i+1} \circ s_i = s_{i+1} \circ s_i \circ s_{i+1} \qquad \text{for all } i \in \{1, 2, \ldots, n-2\} \qquad (310)$$

  (according to Exercise 5.1 **(a)**). Also, it is easy to check that

$$s_i \circ s_j = s_j \circ s_i \qquad \text{for all } i, j \in \{1, 2, \ldots, n-1\} \text{ with } |i-j| > 1. \qquad (311)$$

- An *inversion* of a permutation $\sigma \in S_n$ means a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. The *length* $\ell(\sigma)$ of a permutation $\sigma \in S_n$ is the number of inversions of σ.

- Any two permutations $\sigma \in S_n$ and $\tau \in S_n$ satisfy

$$\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2 \qquad (312)$$

  (by Exercise 5.2 **(b)**) and

$$\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau) \qquad (313)$$

  (by Exercise 5.2 **(c)**).

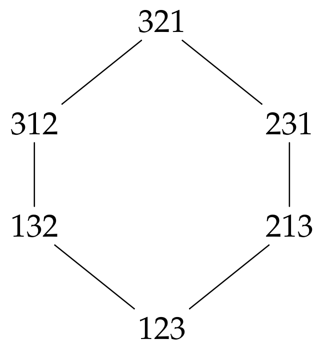
- If $\sigma \in S_n$, then $\ell(\sigma) = \ell(\sigma^{-1})$ (according to Exercise 5.2 **(f)**).

- If $\sigma \in S_n$, then $\ell(\sigma)$ is the smallest $N \in \mathbb{N}$ such that σ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). (This follows from Exercise 5.2 **(g)**.)

By now, we know almost all about the s_k's and about the lengths of permutations that is necessary for studying determinants. ("Almost" because Exercise 5.10 below will also be useful.) I shall now sketch some more advanced properties of these things, partly as a curiosity, partly to further your intuition; none of these properties shall be used further below. The rest of Section 5.5 shall rely on some notions we have not introduced in these notes; in particular, we will use the concepts of undirected graphs ([LeLeMe16, Chapter 12]), directed graphs ([LeLeMe16, Chapter 10]) and (briefly) polytopes (see, e.g., [AigZie14, Chapter 10]).

First, here is a way to visualize lengths of permutations using graph theory:

Fix $n \in \mathbb{N}$. We define the *n-th right Bruhat graph* to be the (undirected) graph whose vertices are the permutations $\sigma \in S_n$, and whose edges are defined as follows: Two vertices $\sigma \in S_n$ and $\tau \in S_n$ are adjacent if and only if there exists a $k \in \{1, 2, \ldots, n-1\}$ such that $\sigma = \tau \circ s_k$. (This condition is clearly symmetric in σ and τ: If $\sigma = \tau \circ s_k$, then $\tau = \sigma \circ s_k$.) For instance, the 3-rd right Bruhat graph looks as follows:



,

where we are writing permutations in one-line notation (and omitting parentheses and commas). The 4-th right Bruhat graph can be seen on Wikipedia.[162]

These graphs have lots of properties. There is a canonical way to direct their edges: The edge between σ and τ is directed towards the vertex with the larger length. (The lengths of σ and τ always differ by 1 if there is an edge between σ and τ.) This way, the n-th right Bruhat graph is an acyclic directed graph. It therefore defines a partially ordered set, called the *right permutohedron order*[163] on S_n, whose elements are the permutations $\sigma \in S_n$ and whose order relation is defined as follows: We have $\sigma \leq \tau$ if and only if there is a directed path from σ to τ in the directed n-th right Bruhat graph. If you know the (combinatorial) notion of a lattice, you might notice that this right permutohedron order is a lattice.

The word "permutohedron" in "permutohedron order" hints at what might be its least expected property: The n-th Bruhat graph can be viewed as the graph formed by the vertices and the edges of a certain polytope in n-dimensional space $\mathbb{R}^n$. This polytope – called the *n-th permutohedron*[164] – is the convex hull of the

[162] Don't omit the word "right" in "right Bruhat graph"; else it means a different graph with more edges.

[163] also known as the *right weak order* or *right weak Bruhat order* (but, again, do not omit the words "right" and "weak")

[164] Some spell it "*permutahedron*" instead. The word is of relatively recent origin (1969).

points $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ for $\sigma \in S_n$. These points are its vertices; however, its vertex $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ corresponds to the vertex σ^{-1} (not σ) of the n-th Bruhat graph. Feel free to roam its Wikipedia page for other (combinatorial and geometric) curiosities.

The notion of a length fits perfectly into this whole picture. For instance, the length $\ell(\sigma)$ of a permutation σ is the smallest length of a path from $\operatorname{id} \in S_n$ to σ on the n-th right Bruhat graph (and this holds no matter whether the graph is considered to be directed or not). For the undirected Bruhat graphs, we have something more general:

Exercise 5.8. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ and $\tau \in S_n$. Show that $\ell\left(\sigma^{-1} \circ \tau\right)$ is the smallest length of a path between σ and τ on the (undirected) n-th right Bruhat graph.

(Recall that the length of a path in a graph is defined as the number of edges along this path.)

How many permutations in S_n have a given length? The number is not easy to compute directly; however, its generating function is nice. (See [LeLeMe16, Chapter 16] for the notion of generating functions.) Namely,

$$\sum_{w \in S_n} q^{\ell(w)} = (1+q)\left(1+q+q^2\right)\cdots\left(1+q+q^2+\cdots+q^{n-1}\right)$$

(where q is an indeterminate). This equality (with q renamed as x) is Corollary 5.53 (which is proven below, in the solution to Exercise 5.18). Another proof can be found in [Stanle11, Corollary 1.3.13] (but notice that Stanley denotes S_n by $\mathfrak{S}_n$, and $\ell(w)$ by $\operatorname{inv}(w)$).

Remark 5.27. Much more can be said. Let me briefly mention (without proof) two other related results.

We can ask ourselves in what different ways a permutation can be written as a composition of N permutations of the form s_k. For instance, the permutation $w_0 \in S_3$ which sends 1, 2 and 3 to 3, 2 and 1, respectively (that is, $w_0 = (3,2,1)$ in one-line notation) can be written as a product of three s_k's in the two forms

$$w_0 = s_1 \circ s_2 \circ s_1, \qquad w_0 = s_2 \circ s_1 \circ s_2, \tag{314}$$

but can also be written as a product of five s_k's (e.g., as $w_0 = s_1 \circ s_2 \circ s_1 \circ s_2 \circ s_2$) or seven s_k's or nine s_k's, etc. Are the different representations of w_0 related?

Clearly, the two representations in (314) are connected to each other by the equality $s_1 \circ s_2 \circ s_1 = s_2 \circ s_1 \circ s_2$, which is a particular case of (310). Also, the representation $w_0 = s_1 \circ s_2 \circ s_1 \circ s_2 \circ s_2$ reduces to $w_0 = s_1 \circ s_2 \circ s_1$ by "cancelling" the two adjacent s_2's at the end (recall that $s_i \circ s_i = s_i^2 = \operatorname{id}$ for every i).

Interestingly, this generalizes. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. A *reduced expression* for σ will mean a representation of σ as a composition of $\ell(\sigma)$ permutations of the form s_k. (As we know, less than $\ell(\sigma)$ such permutations do not suffice; thus the name "reduced".) Then, (one of the many versions of) *Matsumoto's theorem* states that any two reduced expressions of σ can be obtained from one another by a rewriting process, each step of which is either an application of (310) (i.e., you pick an "$s_i \circ s_{i+1} \circ s_i$" in the expression and replace it by "$s_{i+1} \circ s_i \circ s_{i+1}$", or vice versa) or an application of (311) (i.e., you pick an "$s_i \circ s_j$" with $|i-j| > 1$ and replace it by "$s_j \circ s_i$", or vice versa). For instance, for $n = 4$ and $\sigma = (4,3,1,2,5)$ (in one-line notation), the two reduced expressions $\sigma = s_1 \circ s_2 \circ s_3 \circ s_1 \circ s_2$ and $\sigma = s_2 \circ s_3 \circ s_1 \circ s_2 \circ s_3$ can be obtained from one another by the following rewriting process:

$$\begin{aligned} s_1 \circ s_2 \circ \underbrace{s_3 \circ s_1}_{\substack{=s_1 \circ s_3 \\ \text{(by (311))}}} \circ s_2 &= \underbrace{s_1 \circ s_2 \circ s_1}_{\substack{=s_2 \circ s_1 \circ s_2 \\ \text{(by (310))}}} \circ s_3 \circ s_2 = s_2 \circ s_1 \circ \underbrace{s_2 \circ s_3 \circ s_2}_{\substack{=s_3 \circ s_2 \circ s_3 \\ \text{(by (310))}}} \\ &= s_2 \circ \underbrace{s_1 \circ s_3}_{\substack{=s_3 \circ s_1 \\ \text{(by (311))}}} \circ s_2 \circ s_3 = s_2 \circ s_3 \circ s_1 \circ s_2 \circ s_3. \end{aligned}$$

See, e.g., Williamson's thesis [Willia03, Corollary 1.2.3] or Knutson's notes [Knutso12, §2.3] for a proof of this fact. (Knutson, instead of saying that "$\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}$ is a reduced expression for σ", says that "$k_1 k_2 \cdots k_p$ is a reduced word for σ".)

Something subtler holds for "non-reduced" expressions. Namely, if we have a representation of σ as a composition of some number of permutations of the form s_k (not necessarily $\ell(\sigma)$ of them), then we can transform it into a reduced expression by a rewriting process which consists of applications of (310) and (311) as before and also of cancellation steps (i.e., picking an "$s_i \circ s_i$" in the expression and removing it). This follows from [LLPT95, Chapter SYM, Proposition (2.6)][165], and can also easily be derived from [Willia03, Corollary 1.2.3 and Corollary 1.1.6].

This all is stated and proven in greater generality in good books on Coxeter groups, such as [BjoBre05]. We won't need these results in the following, but they are an example of what one can see if one looks at permutations closely.

5.6. More on signs of permutations

In Section 5.3, we have defined the sign $(-1)^\sigma = \operatorname{sign}\sigma = \operatorname{sgn}\sigma$ of a permutation σ. We recall the most important facts about it:

[165] What the authors of [LLPT95] call a "presentation" of a permutation $\sigma \in S_n$ is a finite list $\left(s_{k_1}, s_{k_2}, \ldots, s_{k_p}\right)$ of elements of $\{s_1, s_2, \ldots, s_{n-1}\}$ satisfying $\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}$. What the authors of [LLPT95] call a "minimal presentation" of σ is what we call a reduced expression of σ.

- We have $(-1)^{\sigma} = (-1)^{\ell(\sigma)}$ for every $\sigma \in S_n$. (This is the definition of $(-1)^{\sigma}$.) Thus, for every $\sigma \in S_n$, we have $(-1)^{\sigma} = (-1)^{\ell(\sigma)} \in \{1, -1\}$.

- The permutation $\sigma \in S_n$ is said to be *even* if $(-1)^{\sigma} = 1$, and is said to be *odd* if $(-1)^{\sigma} = -1$.

- The sign of the identity permutation $\mathrm{id} \in S_n$ is $(-1)^{\mathrm{id}} = 1$.

- For every $k \in \{1, 2, \ldots, n-1\}$, the sign of the permutation $s_k \in S_n$ is $(-1)^{s_k} = -1$.

- If σ and τ are two permutations in S_n, then
$$(-1)^{\sigma \circ \tau} = (-1)^{\sigma} \cdot (-1)^{\tau}. \qquad (315)$$

- If $\sigma \in S_n$, then
$$(-1)^{\sigma^{-1}} = (-1)^{\sigma}. \qquad (316)$$

A simple consequence of the above facts is the following proposition:

Proposition 5.28. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Let $\sigma_1, \sigma_2, \ldots, \sigma_k$ be k permutations in S_n. Then,
$$(-1)^{\sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_k} = (-1)^{\sigma_1} \cdot (-1)^{\sigma_2} \cdot \cdots \cdot (-1)^{\sigma_k}. \qquad (317)$$

Proof of Proposition 5.28. Straightforward induction over k. The induction base (i.e., the case when $k = 0$) follows from the fact that $(-1)^{\mathrm{id}} = 1$ (since the composition of 0 permutations is id). The induction step is easily done using (315). □

Let us introduce another notation:

Definition 5.29. Let $n \in \mathbb{N}$. Let i and j be two distinct elements of $\{1, 2, \ldots, n\}$. We let $t_{i,j}$ be the permutation in S_n which switches i with j while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged. Such a permutation is called a *transposition* (and is often denoted by (i, j) in literature; but we prefer not to do so, since it is too similar to one-line notation).

Notice that the permutations $s_1, s_2, \ldots, s_{n-1}$ are transpositions (namely, $s_i = t_{i,i+1}$ for every $i \in \{1, 2, \ldots, n-1\}$), but they are not the only transpositions (when $n \geq 3$).

For the next exercise, we need one further definition, which extends Definition 5.29:

Definition 5.30. Let $n \in \mathbb{N}$. Let i and j be two elements of $\{1, 2, \ldots, n\}$. We define a permutation $t_{i,j} \in S_n$ as follows:

- If $i \neq j$, then the permutation $t_{i,j}$ has already been defined in Definition 5.29.

- If $i = j$, then we define the permutation $t_{i,j}$ to be the identity $\mathrm{id} \in S_n$.

Exercise 5.9. Whenever m is an integer, we shall use the notation $[m]$ for the set $\{1,2,\ldots,m\}$.

Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Prove that there is a unique n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that

$$\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}.$$

Example 5.31. For this example, set $n = 4$, and let $\sigma \in S_4$ be the permutation that sends $1,2,3,4$ to $3,1,4,2$. Then, Exercise 5.9 claims that there is a unique 4-tuple $(i_1, i_2, i_3, i_4) \in [1] \times [2] \times [3] \times [4]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ t_{3,i_3} \circ t_{4,i_4}$.

This 4-tuple can easily be found: it is $(1,1,1,3)$. In fact, we have $\sigma = t_{1,1} \circ t_{2,1} \circ t_{3,1} \circ t_{4,3}$.

Exercise 5.10. Let $n \in \mathbb{N}$. Let i and j be two distinct elements of $\{1,2,\ldots,n\}$.

(a) Find $\ell(t_{i,j})$.

(b) Show that $(-1)^{t_{i,j}} = -1$.

Exercise 5.11. Let $n \in \mathbb{N}$. Let w_0 denote the permutation in S_n which sends each $k \in \{1,2,\ldots,n\}$ to $n+1-k$. Compute $\ell(w_0)$ and $(-1)^{w_0}$.

Exercise 5.12. Let X be a finite set. We want to define the sign of any permutation of X. (We have sketched this definition before (see (295)), but now we shall do it in detail.)

Fix a bijection $\phi : X \to \{1,2,\ldots,n\}$ for some $n \in \mathbb{N}$. (Such a bijection always exists. Indeed, constructing such a bijection is tantamount to writing down a list of all elements of X, with no duplicates.) For every permutation σ of X, set

$$(-1)^{\sigma}_{\phi} = (-1)^{\phi \circ \sigma \circ \phi^{-1}}.$$

Here, the right hand side is well-defined because $\phi \circ \sigma \circ \phi^{-1}$ is a permutation of $\{1,2,\ldots,n\}$.

(a) Prove that $(-1)^{\sigma}_{\phi}$ depends only on the permutation σ of X, but not on the bijection ϕ. (In other words, for a given σ, any two different choices of ϕ will lead to the same $(-1)^{\sigma}_{\phi}$.)

This allows us to define the *sign* of the permutation σ to be $(-1)^{\sigma}_{\phi}$ for any choice of bijection $\phi : X \to \{1,2,\ldots,n\}$. We denote this sign simply by $(-1)^{\sigma}$. (When $X = \{1,2,\ldots,n\}$, then this sign is clearly the same as the sign $(-1)^{\sigma}$ we defined before, because we can pick the bijection $\phi = \mathrm{id}$.)

(b) Show that the permutation $\mathrm{id} : X \to X$ satisfies $(-1)^{\mathrm{id}} = 1$.

(c) Show that $(-1)^{\sigma \circ \tau} = (-1)^{\sigma} \cdot (-1)^{\tau}$ for any two permutations σ and τ of X.

Remark 5.32. A sufficiently pedantic reader might have noticed that the definition of $(-1)^\sigma$ in Exercise 5.12 is not completely kosher. In fact, the set X may be $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$; in this case, σ is an element of S_n, and thus the sign $(-1)^\sigma$ has already been defined in Definition 5.14. Thus, in this case, we are defining the notation $(-1)^\sigma$ a second time in Exercise 5.12. Woe to us if this second definition yields a different number than the first!

Fortunately, it does not. The definition of $(-1)^\sigma$ in Exercise 5.12 does not conflict with the original meaning of $(-1)^\sigma$ as defined in Definition 5.14. Indeed, in order to prove this, we temporarily rename the number $(-1)^\sigma$ defined in Exercise 5.12 as $(-1)^\sigma_{\text{new}}$ (in order to ensure that we don't confuse it with the number $(-1)^\sigma$ defined in Definition 5.14). Now, consider the situation of Exercise 5.12, and assume that $X = \{1, 2, \ldots, n\}$. We must then prove that $(-1)^\sigma_{\text{new}} = (-1)^\sigma$. But the definition of $(-1)^\sigma_{\text{new}}$ in Exercise 5.12 says that $(-1)^\sigma_{\text{new}} = (-1)^\sigma_\phi$, where ϕ is any bijection $X \to \{1, 2, \ldots, n\}$. We can apply this to $\phi = \text{id}$ (because clearly, id is a bijection $X \to \{1, 2, \ldots, n\}$), and thus obtain $(-1)^\sigma_{\text{new}} = (-1)^\sigma_{\text{id}}$. But the definition of $(-1)^\sigma_{\text{id}}$ yields $(-1)^\sigma_{\text{id}} = (-1)^{\text{id} \circ \sigma \circ \text{id}^{-1}} = (-1)^\sigma$ (since $\text{id} \circ \sigma \circ \underbrace{\text{id}^{-1}}_{=\text{id}} = \sigma$). Thus, $(-1)^\sigma_{\text{new}} = (-1)^\sigma_{\text{id}} = (-1)^\sigma$. This is precisely what we wanted to prove. Thus, we have shown that the definition of $(-1)^\sigma$ in Exercise 5.12 does not conflict with the original meaning of $(-1)^\sigma$ as defined in Definition 5.14.

Remark 5.33. Let $n \in \mathbb{N}$. Recall that a *transposition* in S_n means a permutation of the form $t_{i,j}$, where i and j are two distinct elements of $\{1, 2, \ldots, n\}$. Therefore, if τ is a transposition in S_n, then

$$(-1)^\tau = -1. \tag{318}$$

(In fact, if τ is a transposition in S_n, then τ can be written in the form $\tau = t_{i,j}$ for two distinct elements i and j of $\{1, 2, \ldots, n\}$; and therefore, for these two elements i and j, we have $(-1)^\tau = (-1)^{t_{i,j}} = -1$ (according to Exercise 5.10 **(b)**). This proves (318).)

Now, let $\sigma \in S_n$ be any permutation. Assume that σ is written in the form $\sigma = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$ for some transpositions $\tau_1, \tau_2, \ldots, \tau_k$ in S_n. Then,

$$\begin{aligned}(-1)^\sigma = (-1)^{\tau_1 \circ \tau_2 \circ \cdots \circ \tau_k} &= \underbrace{(-1)^{\tau_1}}_{\substack{=-1 \\ \text{(by (318))}}} \cdot \underbrace{(-1)^{\tau_2}}_{\substack{=-1 \\ \text{(by (318))}}} \cdot \cdots \cdot \underbrace{(-1)^{\tau_k}}_{\substack{=-1 \\ \text{(by (318))}}} \\ &\qquad \text{(by (317), applied to } \sigma_i = \tau_i\text{)} \\ &= \underbrace{(-1) \cdot (-1) \cdot \cdots \cdot (-1)}_{k \text{ factors}} = (-1)^k. \end{aligned} \tag{319}$$

Since many permutations can be written as products of transpositions in a simple way, this formula gives a useful method for computing signs.

Remark 5.34. Let $n \in \mathbb{N}$. It is not hard to prove that

$$(-1)^{\sigma} = \prod_{1 \leq i < j \leq n} \frac{\sigma(i) - \sigma(j)}{i - j} \qquad \text{for every } \sigma \in S_n. \tag{320}$$

(Of course, it is no easier to compute $(-1)^{\sigma}$ using this seemingly explicit formula than by counting inversions.)

We shall prove (320) in Exercise 5.13 **(c)**.

Remark 5.35. The sign of a permutation is also called its *signum* or its *signature*. Different authors define the sign of a permutation σ in different ways. Some (e.g., Hefferon in [Heffer17, Chapter Four, Definition 4.7], or Strickland in [Strick13, Definition B.4]) define it as we do; others (e.g., Conrad in [Conrad1] or Hoffman and Kunze in [HofKun71, p. 152]) define it using (319); yet others define it using something called the *cycle decomposition* of a permutation; some even define it using (320), or using a similar ratio of two polynomials. However, it is not hard to check that all of these definitions are equivalent. (We already know that the first two of them are equivalent.)

Exercise 5.13. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$.

(a) If $x_1, x_2, \ldots, x_n$ are n elements of $\mathbb{C}$, then prove that

$$\prod_{1 \leq i < j \leq n} \left(x_{\sigma(i)} - x_{\sigma(j)} \right) = (-1)^{\sigma} \cdot \prod_{1 \leq i < j \leq n} \left(x_i - x_j \right).$$

(b) More generally: For every $(i, j) \in \{1, 2, \ldots, n\}^2$, let $a_{(i,j)}$ be an element of $\mathbb{C}$. Assume that

$$a_{(j,i)} = -a_{(i,j)} \qquad \text{for every } (i, j) \in \{1, 2, \ldots, n\}^2. \tag{321}$$

Prove that

$$\prod_{1 \leq i < j \leq n} a_{(\sigma(i), \sigma(j))} = (-1)^{\sigma} \cdot \prod_{1 \leq i < j \leq n} a_{(i,j)}.$$

(c) Prove (320).

(d) Use Exercise 5.13 **(a)** to give a new solution to Exercise 5.2 **(b)**.

The next exercise relies on the notion of "the list of all elements of S in increasing order (with no repetitions)", where S is a finite set of integers. This notion means exactly what it says; it was rigorously defined in Definition 2.50.

Exercise 5.14. Let $n \in \mathbb{N}$. Let I be a subset of $\{1, 2, \ldots, n\}$. Let $k = |I|$. Let $(a_1, a_2, \ldots, a_k)$ be the list of all elements of I in increasing order (with no repetitions). Let $(b_1, b_2, \ldots, b_{n-k})$ be the list of all elements of $\{1, 2, \ldots, n\} \setminus I$ in increasing order (with no repetitions). Let $\alpha \in S_k$ and $\beta \in S_{n-k}$. Prove the following:

(a) There exists a unique $\sigma \in S_n$ satisfying

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right).$$

Denote this σ by $\sigma_{I,\alpha,\beta}$.

(b) Let $\sum I$ denote the sum of all elements of I. (Thus, $\sum I = \sum_{i \in I} i$.) We have

$$\ell\left(\sigma_{I,\alpha,\beta}\right) = \ell(\alpha) + \ell(\beta) + \sum I - (1 + 2 + \cdots + k)$$

and

$$(-1)^{\sigma_{I,\alpha,\beta}} = (-1)^{\alpha} \cdot (-1)^{\beta} \cdot (-1)^{\sum I - (1+2+\cdots+k)}.$$

(c) Forget that we fixed α and β. We thus have defined an element $\sigma_{I,\alpha,\beta} \in S_n$ for every $\alpha \in S_k$ and every $\beta \in S_{n-k}$. The map

$$S_k \times S_{n-k} \to \{\tau \in S_n \mid \tau(\{1, 2, \ldots, k\}) = I\},$$
$$(\alpha, \beta) \mapsto \sigma_{I,\alpha,\beta}$$

is well-defined and a bijection.

We can define transpositions not only in the symmetric group S_n, but also more generally for arbitrary sets X:

Definition 5.36. Let X be a set. Let i and j be two distinct elements of X. We let $t_{i,j}$ be the permutation of X which switches i with j while leaving all other elements of X unchanged. Such a permutation is called a *transposition* of X.

Clearly, Definition 5.36 is a generalization of Definition 5.29.

Exercise 5.15. Let X be a finite set. Recall that if σ is any permutation of X, then the sign $(-1)^{\sigma}$ of σ is well-defined (by Exercise 5.12). Prove the following:

(a) For any two distinct elements i and j of X, we have $(-1)^{t_{i,j}} = -1$.

(b) Any permutation of X can be written as a composition of finitely many transpositions of X.

(c) Let σ be a permutation of X. Assume that σ can be written as a composition of k transpositions of X. Then, $(-1)^{\sigma} = (-1)^{k}$.

5.7. Cycles

Next, we shall discuss another specific class of permutations: the *cycles*.

Definition 5.37. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$.

Let $k \in \{1, 2, \ldots, n\}$. Let $i_1, i_2, \ldots, i_k$ be k distinct elements of $[n]$. We define $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ to be the permutation in S_n which sends $i_1, i_2, \ldots, i_k$ to $i_2, i_3, \ldots, i_k, i_1,$

respectively, while leaving all other elements of $[n]$ fixed. In other words, we define $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ to be the permutation in S_n given by

$$\left(\operatorname{cyc}_{i_1,i_2,\ldots,i_k}(p) = \begin{cases} i_{j+1}, & \text{if } p = i_j \text{ for some } j \in \{1,2,\ldots,k\}; \\ p, & \text{otherwise} \end{cases} \qquad \text{for every } p \in [n] \right),$$

where i_{k+1} means i_1.

(Again, the notation $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ conceals the parameter n, which will hopefully not cause any confusion.)

A permutation of the form $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ is said to be a *k-cycle* (or sometimes just a *cycle*, or a *cyclic permutation*). Of course, the name stems from the fact that it "cycles" through the elements $i_1, i_2, \ldots, i_k$ (by sending each of them to the next one and the last one back to the first) and leaves all other elements unchanged.

Example 5.38. Let $n \in \mathbb{N}$. The following facts follow easily from Definition 5.37:

(a) For every $i \in \{1,2,\ldots,n\}$, we have $\operatorname{cyc}_i = \operatorname{id}$. In other words, any 1-cycle is the identity permutation id.

(b) If i and j are two distinct elements of $\{1,2,\ldots,n\}$, then $\operatorname{cyc}_{i,j} = t_{i,j}$. (See Definition 5.29 for the definition of $t_{i,j}$.)

(c) If $k \in \{1,2,\ldots,n-1\}$, then $\operatorname{cyc}_{k,k+1} = s_k$.

(d) If $n = 5$, then $\operatorname{cyc}_{2,5,3}$ is the permutation which sends 1 to 1, 2 to 5, 3 to 2, 4 to 4, and 5 to 3. (In other words, it is the permutation which is $(1,5,2,4,3)$ in one-line notation.)

(e) If $k \in \{1,2,\ldots,n\}$, and if $i_1, i_2, \ldots, i_k$ are k pairwise distinct elements of $[n]$, then

$$\operatorname{cyc}_{i_1,i_2,\ldots,i_k} = \operatorname{cyc}_{i_2,i_3,\ldots,i_k,i_1} = \operatorname{cyc}_{i_3,i_4,\ldots,i_k,i_1,i_2} = \cdots = \operatorname{cyc}_{i_k,i_1,i_2,\ldots,i_{k-1}}.$$

(In less formal words: The k-cycle $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ does not change when we cyclically rotate the list $(i_1, i_2, \ldots, i_k)$.)

Remark 5.39. What we called $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ in Definition 5.37 is commonly denoted by $(i_1, i_2, \ldots, i_k)$ in the literature. But this latter notation $(i_1, i_2, \ldots, i_k)$ would clash with one-line notation for permutations (the cycle $\operatorname{cyc}_{1,2,3} \in S_3$ is not the same as the permutation which is $(1,2,3)$ in one-line notation) and also with the standard notation for k-tuples. This is why we prefer to use the notation $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$. (That said, we are not going to use k-cycles very often.)

Any k-cycle is a composition of $k-1$ transpositions, as the following exercise shows:

Exercise 5.16. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$. Let $k \in \{1, 2, \ldots, n\}$. Let $i_1, i_2, \ldots, i_k$ be k distinct elements of $[n]$. Prove that

$$\operatorname{cyc}_{i_1, i_2, \ldots, i_k} = t_{i_1, i_2} \circ t_{i_2, i_3} \circ \cdots \circ t_{i_{k-1}, i_k}.$$

(We are using Definition 5.29 here.)

The following exercise gathers some further properties of cycles. Parts **(a)** and **(d)** and, to a lesser extent, **(b)** are fairly important and you should make sure you know how to solve them. The significantly more difficult part **(c)** is more of a curiosity with an interesting proof (I have not found an application of it so far; skip it if you do not want to spend time on what is essentially a contest problem).

Exercise 5.17. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$. Let $k \in \{1, 2, \ldots, n\}$.

(a) For every $\sigma \in S_n$ and every k distinct elements $i_1, i_2, \ldots, i_k$ of $[n]$, prove that

$$\sigma \circ \operatorname{cyc}_{i_1, i_2, \ldots, i_k} \circ \sigma^{-1} = \operatorname{cyc}_{\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k)}.$$

(b) For every $p \in \{0, 1, \ldots, n-k\}$, prove that

$$\ell\left(\operatorname{cyc}_{p+1, p+2, \ldots, p+k}\right) = k - 1.$$

(c) For every k distinct elements $i_1, i_2, \ldots, i_k$ of $[n]$, prove that

$$\ell\left(\operatorname{cyc}_{i_1, i_2, \ldots, i_k}\right) \geq k - 1.$$

(d) For every k distinct elements $i_1, i_2, \ldots, i_k$ of $[n]$, prove that

$$(-1)^{\operatorname{cyc}_{i_1, i_2, \ldots, i_k}} = (-1)^{k-1}.$$

Remark 5.40. Exercise 5.17 **(d)** shows that every k-cycle in S_n has sign $(-1)^{k-1}$. However, the length of a k-cycle need not be $k-1$. Exercise 5.17 **(c)** shows that this length is always $\geq k-1$, but it can take other values as well. For instance, in S_4, the length of the 3-cycle $\operatorname{cyc}_{1,4,3}$ is 4. (Another example are the transpositions $t_{i,j}$ from Definition 5.29; these are 2-cycles but can have length > 1.)

I don't know a simple way to describe when equality holds in Exercise 5.17 **(c)**. It holds whenever $i_1, i_2, \ldots, i_k$ are consecutive integers (due to Exercise 5.17 **(b)**), but also in some other cases; for example, the 4-cycle $\operatorname{cyc}_{1,3,4,2}$ in S_4 has length 3.

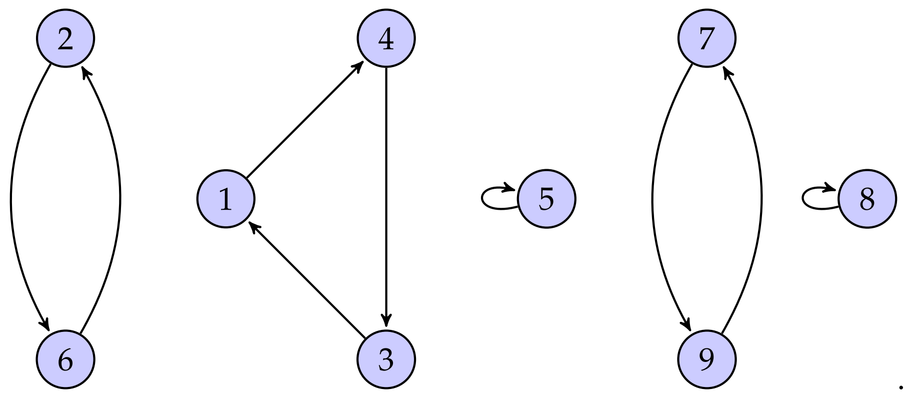
Remark 5.41. The main reason why cycles are useful is that, essentially, every permutation can be "decomposed" into cycles. We shall not use this fact, but since it is generally important, let us briefly explain what it means. (You will probably learn more about it in any standard course on abstract algebra.)

Fix $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$. Two cycles α and β in S_n are said to be *disjoint* if they can be written as $\alpha = \operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ and $\beta = \operatorname{cyc}_{j_1, j_2, \ldots, j_\ell}$ for $k + \ell$ distinct elements $i_1, i_2, \ldots, i_k, j_1, j_2, \ldots, j_\ell$ of $[n]$. For example, the two cycles $\operatorname{cyc}_{1,3}$ and $\operatorname{cyc}_{2,6,7}$ in S_8 are disjoint, but the two cycles $\operatorname{cyc}_{1,4}$ and $\operatorname{cyc}_{2,4}$ are not. It is easy to see that any two disjoint cycles α and β commute (i.e., satisfy $\alpha \circ \beta = \beta \circ \alpha$). Therefore, when you see a composition $\alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_p$ of several pairwise disjoint cycles, you can reorder its factors arbitrarily without changing the result (for example, $\alpha_3 \circ \alpha_1 \circ \alpha_4 \circ \alpha_2 = \alpha_1 \circ \alpha_2 \circ \alpha_3 \circ \alpha_4$ if $p = 4$).

Now, the fact I am talking about says the following: Every permutation in S_n can be written as a composition of several pairwise disjoint cycles. For example, let $n = 9$, and let $\sigma \in S_9$ be the permutation which is written $(4, 6, 1, 3, 5, 2, 9, 8, 7)$ in one-line notation (i.e., we have $\sigma(1) = 4$, $\sigma(2) = 6$, etc.). Then, σ can be written as a composition of several pairwise disjoint cycles as follows:

$$\sigma = \operatorname{cyc}_{1,4,3} \circ \operatorname{cyc}_{7,9} \circ \operatorname{cyc}_{2,6}. \qquad (322)$$

Indeed, here is how such a decomposition can be found: Let us draw a directed graph whose vertices are $1, 2, \ldots, n$, and which has an arc $i \to \sigma(i)$ for every $i \in [n]$. (Thus, it has n arcs altogether; some of them can be loops.) For our permutation $\sigma \in S_9$, this graph looks as follows:



. (323)

Obviously, at each vertex i of this graph, exactly one arc begins (namely, the arc $i \to \sigma(i)$). Moreover, since σ is invertible, it is also clear that at each vertex i of this graph, exactly one arc ends (namely, the arc $\sigma^{-1}(i) \to i$). Due to the way we constructed this graph, it is clear that it completely describes our permutation σ: Namely, if we want to find $\sigma(i)$ for a given $i \in [n]$, we should just locate the vertex i on the graph, and follow the arc that begins at this vertex; the endpoint of this arc will be $\sigma(i)$.

Now, a look at this graph reveals five directed cycles (in the sense of "paths which end at the same vertex at which they begin", not yet in the sense of "cyclic permutations"). The first one passes through the vertices 2 and 6; the second passes through the vertices 3, 1 and 4; the third, through the vertex 5 (it is what is called a "trivial cycle"), and so on. To each of these cycles we can assign a cyclic permutation in S_n: namely, if the cycle passes through the vertices $i_1, i_2, \ldots, i_k$ (in this order, and with no repetitions), then we assign to it the cyclic permutation $\operatorname{cyc}_{i_1,i_2,\ldots,i_k} \in S_n$. The cyclic permutations assigned to all five directed cycles are pairwise disjoint, and their composition is

$$\operatorname{cyc}_{2,6} \circ \operatorname{cyc}_{3,1,4} \circ \operatorname{cyc}_5 \circ \operatorname{cyc}_{7,9} \circ \operatorname{cyc}_8 .$$

But this composition must be σ (because if we apply this composition to an element $i \in [n]$, then we obtain the "next vertex after i" on the directed cycle which passes through i; but due to how we constructed our graph, this "next vertex" will be precisely $\sigma(i)$). Hence, we have

$$\sigma = \operatorname{cyc}_{2,6} \circ \operatorname{cyc}_{3,1,4} \circ \operatorname{cyc}_5 \circ \operatorname{cyc}_{7,9} \circ \operatorname{cyc}_8 . \tag{324}$$

Thus, we have found a way to write σ as a composition of several pairwise disjoint cycles. We can rewrite (and even simplify) this representation a bit: Namely, we can simplify (324) by removing the factors cyc_5 and cyc_8 (because both of these factors equal id); thus we obtain $\sigma = \operatorname{cyc}_{2,6} \circ \operatorname{cyc}_{3,1,4} \circ \operatorname{cyc}_{7,9}$. We can furthermore switch $\operatorname{cyc}_{2,6}$ with $\operatorname{cyc}_{3,1,4}$ (since disjoint cycles commute), therefore obtaining $\sigma = \operatorname{cyc}_{3,1,4} \circ \operatorname{cyc}_{2,6} \circ \operatorname{cyc}_{7,9}$. Next, we can switch $\operatorname{cyc}_{2,6}$ with $\operatorname{cyc}_{7,9}$, obtaining $\sigma = \operatorname{cyc}_{3,1,4} \circ \operatorname{cyc}_{7,9} \circ \operatorname{cyc}_{2,6}$. Finally, we can rewrite $\operatorname{cyc}_{3,1,4}$ as $\operatorname{cyc}_{1,4,3}$, and we obtain (322).

In general, for every $n \in \mathbb{N}$, every permutation $\sigma \in S_n$ can be represented as a composition of several pairwise disjoint cycles (which can be found by drawing a directed graph as in our example above). This representation is not literally unique, because we can modify it by:

- adding or removing trivial factors (i.e., factors of the form $\operatorname{cyc}_i = \mathrm{id}$);
- switching different cycles;
- rewriting $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ as $\operatorname{cyc}_{i_2,i_3,\ldots,i_k,i_1}$.

However, it is unique **up to these modifications**; in other words, any two representations of σ as a composition of several pairwise disjoint cycles can be transformed into one another by such modifications.

The proofs of all these statements are fairly easy. (One does have to check certain things, e.g., that the directed graph really consists of disjoint directed cycles. For a complete proof, see [Goodma15, Theorem 1.5.3] or various other texts on algebra.)

Representing a permutation $\sigma \in S_n$ as a composition of several pairwise disjoint cycles can be done very quickly, and thus gives a quick way to find $(-1)^\sigma$ (because Exercise 5.17 **(d)** tells us how to find the sign of a k-cycle). This is significantly faster than counting inversions of σ.

5.8. The Lehmer code

In this short section, we shall introduce the *Lehmer code* of a permutation. Throughout Section 5.8, we will use the following notations:

Definition 5.42. **(a)** Whenever m is an integer, we shall use the notation $[m]$ for the set $\{1, 2, \ldots, m\}$. (This is an empty set when $m \leq 0$.)

(b) Whenever m is an integer, we shall use the notation $[m]_0$ for the set $\{0, 1, \ldots, m\}$. (This is an empty set when $m < 0$.)

Definition 5.43. Let $n \in \mathbb{N}$. We consider n to be fixed throughout Section 5.8.

Let H denote the set $[n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0$.

Definition 5.44. Let $\sigma \in S_n$ and $i \in [n]$. Then, $\ell_i(\sigma)$ shall denote the number of all $j \in \{i+1, i+2, \ldots, n\}$ such that $\sigma(i) > \sigma(j)$.

Example 5.45. For this example, set $n = 5$, and let $\sigma \in S_5$ be the permutation that sends $1, 2, 3, 4, 5$ to $4, 3, 2, 1, 5$. Then, $\ell_2(\sigma)$ is the number of all $j \in \{3, 4, 5\}$ such that $\sigma(2) > \sigma(j)$. These j are 3 and 4 (because $\sigma(2) > \sigma(3)$ and $\sigma(2) > \sigma(4)$ but not $\sigma(2) > \sigma(5)$); therefore, $\ell_2(\sigma) = 2$. Similarly, $\ell_1(\sigma) = 3$, $\ell_3(\sigma) = 1$, $\ell_4(\sigma) = 0$ and $\ell_5(\sigma) = 0$.

The following two facts are almost trivial:[166]

Proposition 5.46. Let $\sigma \in S_n$. Then, $\ell(\sigma) = \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma)$.

Proposition 5.47. Let $\sigma \in S_n$. Then, $(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \in H$.

The following simple lemma gives two equivalent ways to define $\ell_i(\sigma)$ for $\sigma \in S_n$ and $i \in [n]$:

Lemma 5.48. Let $\sigma \in S_n$ and $i \in [n]$. Then:

(a) We have $\ell_i(\sigma) = |[\sigma(i) - 1] \setminus \sigma([i])|$.

(b) We have $\ell_i(\sigma) = |[\sigma(i) - 1] \setminus \sigma([i-1])|$.

(c) We have $\sigma(i) \leq i + \ell_i(\sigma)$.

Before we state the next proposition, we introduce another notation:

Definition 5.49. Let $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ be two n-tuples of integers. We say that $(a_1, a_2, \ldots, a_n) <_{\text{lex}} (b_1, b_2, \ldots, b_n)$ if and only if there exists some $k \in [n]$ such that $a_k \neq b_k$, and the **smallest** such k satisfies $a_k < b_k$.

[166] See Exercise 5.18 below for the proofs of all the following results.

For example, $(4,1,2,5) <_{\text{lex}} (4,1,3,0)$ and $(1,1,0,1) <_{\text{lex}} (2,0,0,0)$. The relation $<_{\text{lex}}$ is usually pronounced "is lexicographically smaller than"; the word "lexicographic" comes from the idea that if numbers were letters, then a "word" $a_1 a_2 \cdots a_n$ would appear earlier in a dictionary than $b_1 b_2 \cdots b_n$ if and only if $(a_1, a_2, \ldots, a_n) <_{\text{lex}} (b_1, b_2, \ldots, b_n)$.

Proposition 5.50. Let $\sigma \in S_n$ and $\tau \in S_n$ be such that

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) <_{\text{lex}} (\tau(1), \tau(2), \ldots, \tau(n)).$$

Then,

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) <_{\text{lex}} (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau)).$$

We can now define the Lehmer code:

Definition 5.51. Define the map $L : S_n \to H$ by

$$(L(\sigma) = (\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \qquad \text{for each } \sigma \in S_n).$$

(This is well-defined because of Proposition 5.47.)

If $\sigma \in S_n$ is any permutation, then $L(\sigma) = (\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma))$ is called the *Lehmer code* of σ.

Theorem 5.52. The map $L : S_n \to H$ is a bijection.

Using this theorem and Proposition 5.46, we can easily show the following:

Corollary 5.53. We have

$$\sum_{w \in S_n} x^{\ell(w)} = (1+x)\left(1+x+x^2\right)\cdots\left(1+x+x^2+\cdots+x^{n-1}\right)$$

(an equality between polynomials in x). (The right hand side of this equality should be understood as the empty product when $n \leq 1$.)

Exercise 5.18. Prove Proposition 5.46, Proposition 5.47, Lemma 5.48, Proposition 5.50, Theorem 5.52 and Corollary 5.53.

See [Manive01, §2.1] and [Kerber99, §11.3] for further properties of permutations related to the Lehmer code. (In particular, [Manive01, proof of Proposition 2.1.2] and [Kerber99, Corollary 11.3.5] give two different ways of reconstructing a permutation from its Lehmer code; moreover, [Kerber99, Corollary 11.3.5] shows how the Lehmer code of a permutation $\sigma \in S_n$ leads to a specific representation of σ as a product of some of the $s_1, s_2, \ldots, s_{n-1}$.)

Exercise 5.19. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ and $\tau \in S_n$. We shall use the notation from Definition 3.48.

(a) Prove that each $i \in [n]$ satisfies

$$\begin{aligned} & \ell_{\tau(i)}(\sigma) + \ell_i(\tau) - \ell_i(\sigma \circ \tau) \\ & = \sum_{j \in [n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\ & \qquad + \sum_{j \in [n]} [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] . \end{aligned}$$

(b) Prove that

$$\ell(\sigma) + \ell(\tau) - \ell(\sigma \circ \tau) = 2 \sum_{i \in [n]} \sum_{j \in [n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] .$$

(c) Give a new solution to Exercise 5.2 **(a)**.
(d) Give a new solution to Exercise 5.2 **(b)**.
(e) Give a new solution to Exercise 5.2 **(c)**.

Exercise 5.20. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Let i and j be two elements of $[n]$ such that $i < j$ and $\sigma(i) > \sigma(j)$. Let Q be the set of all $k \in \{i+1, i+2, \ldots, j-1\}$ satisfying $\sigma(i) > \sigma(k) > \sigma(j)$. Prove that

$$\ell\left(\sigma \circ t_{i,j}\right) = \ell(\sigma) - 2|Q| - 1.$$

The following exercise shows an explicit way of expressing every permutation $\sigma \in S_n$ as a product of $\ell(\sigma)$ many simple transpositions (i.e., transpositions of the form s_i with $i \in \{1, 2, \ldots, n-1\}$):

Exercise 5.21. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. If u and v are any two elements of $[n]$ such that $u \leq v$, then we define a permutation $c_{u,v} \in S_n$ by

$$c_{u,v} = \operatorname{cyc}_{v,v-1,v-2,\ldots,u} .$$

For each $i \in [n]$, we define a permutation $a_i \in S_n$ by

$$a_i = c_{i,i+\ell_i(\sigma)}.$$

(a) Prove that a_i is well-defined for each $i \in [n]$.
(b) Prove that each $i \in [n]$ satisfies $a_i = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$, where $i' = i + \ell_i(\sigma)$.
(c) Prove that $\sigma = a_1 \circ a_2 \circ \cdots \circ a_n$.
(d) Solve Exercise 5.2 **(e)** again.
(e) Solve Exercise 5.1 **(c)** again.

5.9. Extending permutations

In this short section, we shall discuss a simple yet useful concept: that of extending a permutation of a set Y to a larger set X (where "larger" means that $Y \subseteq X$). The following notations will be used throughout this section:

Definition 5.54. Let X be a set. Then, S_X denotes the set of all permutations of X.

Definition 5.55. Let X be a set. Let Y be a subset of X. For every map $\sigma : Y \to Y$, we define a map $\sigma^{(Y\to X)} : X \to X$ by

$$\left(\sigma^{(Y\to X)}(x) = \begin{cases} \sigma(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \text{for every } x \in X \right).$$

(This map $\sigma^{(Y\to X)}$ is indeed well-defined, according to Proposition 5.56 below.)

The latter of these two definitions relies on the following lemma:

Proposition 5.56. Let X be a set. Let Y be a subset of X. Let $\sigma : Y \to Y$ be a map. Then, the map $\sigma^{(Y\to X)}$ in Definition 5.55 is well-defined.

Proposition 5.56 is easy to prove; its proof is part of Exercise 5.22 further below.

The idea behind the definition of $\sigma^{(Y\to X)}$ in Definition 5.55 is simple: $\sigma^{(Y\to X)}$ is just the most straightforward way of extending $\sigma : Y \to Y$ to a map from X to X (namely, by letting it keep every element of $X \setminus Y$ unchanged).

Example 5.57. **(a)** If $X = \{1,2,3,4,5,6,7\}$ and $Y = \{1,2,3,4\}$, and if $\sigma \in S_Y = S_4$ is the permutation whose one-line notation is $(4,1,3,2)$, then $\sigma^{(Y\to X)} \in S_X = S_7$ is the permutation whose one-line notation is $(4,1,3,2,5,6,7)$.

(b) More generally, if $X = \{1,2,\ldots,n\}$ and $Y = \{1,2,\ldots,m\}$ for two nonnegative integers n and m satisfying $n \geq m$, and if $\sigma \in S_Y = S_m$ is any permutation, then the permutation $\sigma^{(Y\to X)} \in S_X = S_n$ has one-line notation $(\sigma(1), \sigma(2), \ldots, \sigma(m), m+1, m+2, \ldots, n)$.

(c) If $X = \{1,2,3,\ldots\}$ and $Y = \{1,2,\ldots,n\}$ for some $n \in \mathbb{N}$, and if $\sigma \in S_Y = S_n$ is any permutation, then the permutation $\sigma^{(Y\to X)} \in S_X = S_\infty$ is precisely the permutation $\sigma_{(\infty)}$ defined in Remark 5.26.

Here are some further properties of the operation that transforms σ into $\sigma^{(Y\to X)}$:

Proposition 5.58. Let X be a set. Let Y be a subset of X.

(a) If $\alpha : Y \to Y$ and $\beta : Y \to Y$ are two maps, then

$$(\alpha \circ \beta)^{(Y\to X)} = \alpha^{(Y\to X)} \circ \beta^{(Y\to X)}.$$

(b) The map $\operatorname{id}_Y : Y \to Y$ satisfies

$$(\operatorname{id}_Y)^{(Y \to X)} = \operatorname{id}_X .$$

(c) Every permutation $\sigma \in S_Y$ satisfies $\sigma^{(Y \to X)} \in S_X$ and

$$\left(\sigma^{-1}\right)^{(Y \to X)} = \left(\sigma^{(Y \to X)}\right)^{-1}.$$

(d) We have

$$\left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\} = \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}.$$

(e) The map

$$S_Y \to \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\},$$
$$\delta \mapsto \delta^{(Y \to X)}$$

is well-defined and bijective.

Proposition 5.59. Let X be a set. Let Y be a subset of X. Let Z be a subset of Y. Let $\sigma : Z \to Z$ be any map. Then,

$$\left(\sigma^{(Z \to Y)}\right)^{(Y \to X)} = \sigma^{(Z \to X)}.$$

Proposition 5.60. Let X be a set. Let Y be a subset of X. Let $\alpha : Y \to Y$ be a map. Let $\beta : X \setminus Y \to X \setminus Y$ be a map. Then,

$$\alpha^{(Y \to X)} \circ \beta^{(X \setminus Y \to X)} = \beta^{(X \setminus Y \to X)} \circ \alpha^{(Y \to X)}.$$

The above propositions are fairly straightforward; again, see Exercise 5.22 for their proofs. Interestingly, we can use these simple facts to prove the following nontrivial theorem:

Theorem 5.61. Let X be a finite set. Let $\pi \in S_X$. Then, there exists a $\sigma \in S_X$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$.

In fact, we can show (by induction on $|X|$) the following more general fact:

Proposition 5.62. Let X be a finite set. Let $x \in X$. Let $\pi \in S_X$. Then, there exists a $\sigma \in \left\{\delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}}\right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$.

Exercise 5.22. Prove Proposition 5.56, Proposition 5.58, Proposition 5.59, Proposition 5.60, Proposition 5.62 and Theorem 5.61.

Theorem 5.61 is a known fact, and it is commonly obtained as part of the study of conjugacy in symmetric groups. If π_1 and π_2 are two permutations of a set X, then π_1 is said to be *conjugate* to π_2 if and only if there exists some $\sigma \in S_X$ such that $\sigma \circ \pi_1 \circ \sigma^{-1} = \pi_2$. Thus, Theorem 5.61 says that every permutation π of a finite set X is conjugate to its inverse π^{-1}. Standard proofs of this theorem[167] tend to derive it from the fact that two permutations π_1 and π_2 of a finite set X are conjugate to one another[168] if and only if they have the same "cycle type" (see [Conrad3, Theorem 5.7] for what this means and for a proof).

5.10. Additional exercises

Permutations and symmetric groups are a staple of combinatorics; there are countless results involving them. For an example, Bóna's book [Bona12], as well as significant parts of Stanley's [Stanle11] and [Stanle01] are devoted to them. In this section, I shall only give a haphazard selection of exercises, which are not relevant to the rest of these notes (thus can be skipped at will). I am not planning to provide solutions for all of them.

Exercise 5.23. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Let $a_1, a_2, \ldots, a_n$ be any n numbers. (Here, "number" means "real number" or "complex number" or "rational number", as you prefer; this makes no difference.) Prove that

$$\sum_{\substack{1 \leq i < j \leq n; \\ \sigma(i) > \sigma(j)}} \left(a_j - a_i\right) = \sum_{i=1}^{n} a_i \left(i - \sigma\left(i\right)\right).$$

[Here, the summation sign "$\sum\limits_{\substack{1 \leq i < j \leq n; \\ \sigma(i) > \sigma(j)}}$" means "$\sum\limits_{\substack{(i,j) \in \{1,2,\ldots,n\}^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}}$"; this is a sum over all inversions of σ.]

Exercise 5.24. Let $n \in \mathbb{N}$. Let $\pi \in S_n$.

(a) Prove that

$$\sum_{\substack{1 \leq i < j \leq n; \\ \pi(i) > \pi(j)}} \left(\pi\left(j\right) - \pi\left(i\right)\right) = \sum_{\substack{1 \leq i < j \leq n; \\ \pi(i) > \pi(j)}} \left(i - j\right).$$

---

[167] which, incidentally, also holds for infinite sets X, provided that one believes in the Axiom of Choice

[168] It is easy to see that being conjugate is a symmetric relation: If a permutation π_1 is conjugate to a permutation π_2, then π_2 is, in turn, conjugate to π_1.

[Here, the summation sign " $\sum\limits_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}}$ " means " $\sum\limits_{\substack{(i,j)\in\{1,2,\ldots,n\}^2;\\ i<j \text{ and } \pi(i)>\pi(j)}}$ "; this is a sum over all inversions of π.]

(b) Prove that

$$\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}} \left(\pi\left(j\right)-\pi\left(i\right)\right) = \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}} \left(j-i\right).$$

[Here, the summation sign " $\sum\limits_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}}$ " means " $\sum\limits_{\substack{(i,j)\in\{1,2,\ldots,n\}^2;\\ i<j \text{ and } \pi(i)<\pi(j)}}$ ".]

Exercise 5.24 is [SacUlf11, Proposition 2.4].

Exercise 5.25. Whenever m is an integer, we shall use the notation $[m]$ for the set $\{1,2,\ldots,m\}$. Also, recall Definition 5.30.

Let $n\in\mathbb{N}$. Let $\sigma\in S_n$. Exercise 5.9 shows that there is a unique n-tuple $(i_1,i_2,\ldots,i_n)\in[1]\times[2]\times\cdots\times[n]$ such that

$$\sigma = t_{1,i_1}\circ t_{2,i_2}\circ\cdots\circ t_{n,i_n}.$$

Consider this $(i_1,i_2,\ldots,i_n)$.

For each $k\in\{0,1,\ldots,n\}$, we define a permutation $\sigma_k\in S_n$ by $\sigma_k = t_{1,i_1}\circ t_{2,i_2}\circ\cdots\circ t_{k,i_k}$.

For each $k\in[n]$, we let $m_k=\sigma_k\left(k\right)$.

(a) Prove that $\sigma_k\left(i\right)=i$ for each $i\in[n]$ and each $k\in\{0,1,\ldots,i-1\}$.

(b) Prove that $m_k\in[k]$ for all $k\in[n]$.

(c) Prove that $\sigma_k\left(i_k\right)=k$ for all $k\in[n]$.

(d) Prove that $\sigma_k=t_{k,m_k}\circ\sigma_{k-1}$ for all $k\in[n]$.

(e) Show that $\sigma^{-1}=t_{1,m_1}\circ t_{2,m_2}\circ\cdots\circ t_{n,m_n}$.

(f) Let $x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_n$ be any $2n$ numbers (e.g., rational numbers or real numbers or complex numbers). Prove that

$$\sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k y_{\sigma(k)} = \sum_{k=1}^{n}\left(x_{i_k}-x_k\right)\left(y_{m_k}-y_k\right).$$

(g) Now assume that the numbers $x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_n$ are real and satisfy $x_1\geq x_2\geq\cdots\geq x_n$ and $y_1\geq y_2\geq\cdots\geq y_n$. Conclude that

$$\sum_{k=1}^{n} x_k y_k \geq \sum_{k=1}^{n} x_k y_{\sigma(k)}.$$

Remark 5.63. The claim of Exercise 5.25 **(g)** is known as the *rearrangement inequality*. It has several simple proofs (see, e.g., its Wikipedia page); the approach suggested by Exercise 5.25 is probably the most complicated, but it has the advantage of giving an "explicit" formula for the difference between the two sides (in Exercise 5.25 **(f)**).

Exercise 5.26. Let $n \in \mathbb{N}$. Let $d = \operatorname{lcm}(1, 2, \ldots, n)$. (Here, "lcm" stands for the least common multiple of several integers: Thus, $\operatorname{lcm}(1, 2, \ldots, n)$ is the smallest positive integer that is divisible by $1, 2, \ldots, n$.)

(a) Show that $\pi^d = \operatorname{id}$ for every $\pi \in S_n$.

(b) Let k be an integer such that every $\pi \in S_n$ satisfies $\pi^k = \operatorname{id}$. Show that $d \mid k$.

Exercise 5.27. Let U and V be two finite sets. Let σ be a permutation of U. Let τ be a permutation of V. We define a permutation $\sigma \times \tau$ of the set $U \times V$ by setting

$$(\sigma \times \tau)(a, b) = (\sigma(a), \tau(b)) \qquad \text{for every } (a, b) \in U \times V.$$

(a) Prove that $\sigma \times \tau$ is a well-defined permutation.

(b) Prove that $\sigma \times \tau = (\sigma \times \operatorname{id}_V) \circ (\operatorname{id}_U \times \tau)$.

(c) Prove that $(-1)^{\sigma \times \tau} = \left((-1)^{\sigma}\right)^{|V|} \left((-1)^{\tau}\right)^{|U|}$. (See Exercise 5.12 for the definition of the signs $(-1)^{\sigma \times \tau}$, $(-1)^{\sigma}$ and $(-1)^{\tau}$ appearing here.)

Exercise 5.28. Let $n \in \mathbb{N}$. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. For each $\sigma \in S_n$, define an integer $h(\sigma)$ by

$$h(\sigma) = \sum_{i \in [n]} |\sigma(i) - i|.$$

Let $\sigma \in S_n$.

(a) Prove that

$$h(\sigma) = 2 \sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) = 2 \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i)).$$

(b) Prove that $h(\sigma \circ \tau) \leq h(\sigma) + h(\tau)$ for any $\tau \in S_n$.

(c) Prove that $h(s_k \circ \sigma) \leq h(\sigma) + 2$ for each $k \in \{1, 2, \ldots, n-1\}$.

(d) Prove that

$$h(\sigma)/2 \leq \ell(\sigma) \leq h(\sigma).$$

[**Hint:** The second inequality in part **(d)** is tricky. One way to proceed is by classifying all inversions (i, j) of σ into two types: *Type-I inversions* are those that satisfy $\sigma(i) < j$, whereas *Type-II inversions* are those that satisfy $\sigma(i) \geq j$. Prove that the number of Type-I inversions is $\leq \sum\limits_{\substack{j \in [n]; \\ \sigma(j) < j}} (j - \sigma(j) - 1) \leq \sum\limits_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i))$, whereas the number of Type-II inversions is $\leq \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i)$. Add these together to obtain an upper bound on $\ell(\sigma)$.]

Exercise 5.28 is a result of Diaconis and Graham [DiaGra77, (3.5)][169]. The integer $h(\sigma)$ defined in Exercise 5.28 is called *Spearman's disarray* or *total displacement* of σ. A related concept (the depth of a permutation) has been studied by Petersen and Tenner [PetTen14].

The next two exercises concern the inversions of a permutation. They use the following definition:

Definition 5.64. Let $n \in \mathbb{N}$. For every $\sigma \in S_n$, we let $\operatorname{Inv}\sigma$ denote the set of all inversions of σ.

Exercise 5.2 **(c)** shows that any $n \in \mathbb{N}$ and any two permutations σ and τ in S_n satisfy the inequality $\ell(\sigma\circ\tau) \leq \ell(\sigma) + \ell(\tau)$. In the following exercise, we will see when this inequality becomes an equality:

Exercise 5.29. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ and $\tau \in S_n$.

(a) Prove that $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$ holds if and only if $\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)$.

(b) Prove that $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$ holds if and only if $\operatorname{Inv}\left(\sigma^{-1}\right) \subseteq \operatorname{Inv}\left(\tau^{-1}\circ\sigma^{-1}\right)$.

(c) Prove that $\operatorname{Inv}\sigma \subseteq \operatorname{Inv}\tau$ holds if and only if $\ell(\tau) = \ell\left(\tau\circ\sigma^{-1}\right) + \ell(\sigma)$.

(d) Prove that if $\operatorname{Inv}\sigma = \operatorname{Inv}\tau$, then $\sigma = \tau$.

(e) Prove that $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$ holds if and only if $(\operatorname{Inv}\sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$.

Exercise 5.29 **(d)** shows that if two permutations in S_n have the same set of inversions, then they are equal. In other words, a permutation in S_n is uniquely determined by its set of inversions. The next exercise shows what set of inversions a permutation can have:

Exercise 5.30. Let $n \in \mathbb{N}$. Let $G = \left\{(i,j) \in \mathbb{Z}^2 \mid 1 \leq i < j \leq n\right\}$.

A subset U of G is said to be *transitive* if every $a, b, c \in \{1, 2, \ldots, n\}$ satisfying $(a,b) \in U$ and $(b,c) \in U$ also satisfy $(a,c) \in U$.

A subset U of G is said to be *inversive* if there exists a $\sigma \in S_n$ such that $U = \operatorname{Inv}\sigma$.

Let U be a subset of G. Prove that U is inversive if and only if both U and $G \setminus U$ are transitive.

[169] Note that their notations are different; what they call $I(\pi)$ and $D(\pi)$ would be called $\ell(\pi)$ and $h(\pi)$ (respectively) in my terminology.

6. An introduction to determinants

In this chapter, we will define and study determinants in a combinatorial way (in the spirit of Hefferon's book [Heffer17], Gill Williamson's notes [Willia18, Chapter 3], Laue's notes [Laue15] and Zeilberger's paper [Zeilbe85][170]). Nowadays, students usually learn about determinants in the context of linear algebra, after having made the acquaintance of vector spaces, matrices, linear transformations, Gaussian elimination etc.; this approach to determinants (which I like to call the "linear-algebraic approach") has certain advantages and certain disadvantages compared to our combinatorial approach[171].

We shall study determinants of matrices over *commutative rings*.[172] First, let us

---

[170]My notes differ from these sources in the following:

- Hefferon's book [Heffer17] is an introductory textbook for a first course in Linear Algebra, and so treats rather little of the theory of determinants (far less than what we do). It is, however, a good introduction into the "other part" of linear algebra (i.e., the theory of vector spaces and linear maps), and puts determinants into the context of that other part, which makes some of their properties appear less mysterious. (Like many introductory textbooks, it only discusses matrices over fields, not over commutative rings; it also uses more handwaving in the proofs.)
- Zeilberger's paper [Zeilbe85] mostly proves advanced results (apart from its Section 5, which proves our Theorem 6.23). I would recommend reading it after reading this chapter.
- Laue's notes [Laue15] are a brief introduction to determinants that prove the main results in just 14 pages (although at the cost of terser writing and stronger assumptions on the reader's preknowledge). If you read these notes, make sure to pay attention to the "Prerequisites and some Terminology" section, as it explains the (unusual) notations used in these notes.
- Gill Williamson's [Willia18, Chapter 3] probably comes the closest to what I am doing below (and is highly recommended, not least because it goes much further into various interesting directions!). My notes are more elementary and more detailed in what they do.

Other references treating determinants in a combinatorial way are Day's [Day16, Chapter 6], Herstein's [Herstei75, §6.9], Strickland's [Strick13, §12 and Appendix B], Mate's [Mate14], Walker's [Walker87, §5.4], and Pinkham's [Pinkha15, Chapter 11] (but they all limit themselves to the basics).

[171]Its main advantage is that it gives more motivation and context. However, the other (combinatorial) approach requires less preknowledge and involves fewer technical subtleties (for example, it defines the determinant directly by an explicit formula, while the linear-algebraic approach defines it implicitly by a list of conditions which happen to determine it uniquely), which is why I have chosen it.

Examples of texts that introduce determinants via the linear-algebraic approach are [BirMac99, Chapter IX], [GalQua18, Chapter 7], [Goodma15, §8.3], [HofKun71, Chapter 5] and [Hunger03, §VII.3].

Artin, in [Artin10, Chapter 1], takes a particularly quick approach to determinants over a field (although it is quick at the cost of generality: for example, the proof he gives for [Artin10, Theorem 1.4.9] does not generalize to matrices over commutative rings). Axler's [Axler15, Chapter 10B] gives a singularly horrible treatment of determinants – defining them only for real and complex matrices and in a way that utterly hides their combinatorial structure.

[172]This is a rather general setup, which includes determinants of matrices with real entries, of matrices with complex entries, of matrices with polynomial entries, and many other situations.

define what these words ("commutative ring", "matrix" and "determinant") mean.

6.1. Commutative rings

First, let us define the concept of commutative rings, and give some examples for it.

Definition 6.1. If $\mathbb{K}$ is a set, then a *binary operation* on $\mathbb{K}$ means a map from $\mathbb{K} \times \mathbb{K}$ to $\mathbb{K}$. (In other words, it means a function which takes two elements of $\mathbb{K}$ as input, and returns an element of $\mathbb{K}$ as output.) For instance, the map from $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z}$ which sends every pair $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ to $3a - b$ is a binary operation on $\mathbb{Z}$.

Sometimes, a binary operation f on a set $\mathbb{K}$ will be *written infix*. This means that the image of $(a, b) \in \mathbb{K} \times \mathbb{K}$ under f will be denoted by afb instead of $f(a, b)$. For instance, the binary operation $+$ on the set $\mathbb{Z}$ (which sends a pair (a, b) of integers to their sum $a + b$) is commonly written infix, because one writes $a + b$ and not $+(a, b)$ for the sum of a and b.

Definition 6.2. A *commutative ring* means a set $\mathbb{K}$ endowed with

- two binary operations called "addition" and "multiplication", and denoted by $+$ and $\cdot$, respectively, and both written infix[173], and
- two elements called $0_{\mathbb{K}}$ and $1_{\mathbb{K}}$

such that the following axioms are satisfied:

- *Commutativity of addition:* We have $a + b = b + a$ for all $a \in \mathbb{K}$ and $b \in \mathbb{K}$.
- *Commutativity of multiplication:* We have $ab = ba$ for all $a \in \mathbb{K}$ and $b \in \mathbb{K}$. Here and in the following, the expression "ab" is shorthand for "$a \cdot b$" (as is usual for products of numbers).
- *Associativity of addition:* We have $a + (b + c) = (a + b) + c$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$.
- *Associativity of multiplication:* We have $a(bc) = (ab)c$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$.
- *Neutrality of* 0*:* We have $a + 0_{\mathbb{K}} = 0_{\mathbb{K}} + a = a$ for all $a \in \mathbb{K}$.
- *Existence of additive inverses:* For every $a \in \mathbb{K}$, there exists an element $a' \in \mathbb{K}$ such that $a + a' = a' + a = 0_{\mathbb{K}}$. This a' is commonly denoted by $-a$ and called the *additive inverse* of a. (It is easy to check that it is unique.)

One benefit of working combinatorially is that studying determinants in this general setup is no more difficult than studying them in more restricted settings.

- *Unitality (a.k.a. neutrality of* 1*):* We have $1_{\mathbb{K}} a = a 1_{\mathbb{K}} = a$ for all $a \in \mathbb{K}$.
- *Annihilation:* We have $0_{\mathbb{K}} a = a 0_{\mathbb{K}} = 0_{\mathbb{K}}$ for all $a \in \mathbb{K}$.
- *Distributivity:* We have $a(b+c) = ab + ac$ and $(a+b)c = ac + bc$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$. Here and in the following, we are following the usual convention ("PEMDAS") that multiplication-like operations have higher precedence than addition-like operations; thus, the expression "$ab + ac$" must be understood as "$(ab) + (ac)$" (and not, for example, as "$a(b+a)c$").

(Some of these axioms are redundant, in the sense that they can be derived from others. For instance, the equality $(a+b)c = ac + bc$ can be derived from the axiom $a(b+c) = ab + ac$ using commutativity of multiplication. Also, annihilation follows from the other axioms[174]. The reasons why we have chosen these axioms and not fewer (or more, or others) are somewhat a matter of taste. For example, I like to explicitly require annihilation, because it is an important axiom in the definition of a *semiring*, where it no longer follows from the others.)

Definition 6.3. As we have seen in Definition 6.2, a commutative ring consists of a set $\mathbb{K}$, two binary operations on this set named $+$ and $\cdot$, and two elements of this set named 0 and 1. Thus, formally speaking, we should encode a commutative ring as the 5-tuple $(\mathbb{K}, +, \cdot, 0_{\mathbb{K}}, 1_{\mathbb{K}})$. Sometimes we will actually do so; but most of the time, we will refer to the commutative ring just as the "commutative ring $\mathbb{K}$", hoping that the other four entries of the 5-tuple (namely, $+$, $\cdot$, $0_{\mathbb{K}}$ and $1_{\mathbb{K}}$) are clear from the context. This kind of abbreviation is commonplace in mathematics; it is called "*pars pro toto*" (because we are referring to a large structure by the same symbol as for a small part of it, and hoping that the rest can be inferred from the context). It is an example of what is called "abuse of notation".

The elements $0_{\mathbb{K}}$ and $1_{\mathbb{K}}$ of a commutative ring $\mathbb{K}$ are called the *zero* and the *unity*[175] of $\mathbb{K}$. They are usually denoted by 0 and 1 (without the subscript $\mathbb{K}$)

[173] i.e., we write $a + b$ for the image of $(a, b) \in \mathbb{K} \times \mathbb{K}$ under the binary operation called "addition", and we write $a \cdot b$ for the image of $(a, b) \in \mathbb{K} \times \mathbb{K}$ under the binary operation called "multiplication"

[174] In fact, let $a \in \mathbb{K}$. Distributivity yields $(0_{\mathbb{K}} + 0_{\mathbb{K}}) a = 0_{\mathbb{K}} a + 0_{\mathbb{K}} a$, so that $0_{\mathbb{K}} a + 0_{\mathbb{K}} a = \underbrace{(0_{\mathbb{K}} + 0_{\mathbb{K}})}_{\substack{= 0_{\mathbb{K}} \\ \text{(by neutrality of } 0_{\mathbb{K}})}} a = 0_{\mathbb{K}} a$. Adding $-(0_{\mathbb{K}} a)$ on the left, we obtain $-(0_{\mathbb{K}} a) + (0_{\mathbb{K}} a + 0_{\mathbb{K}} a) = -(0_{\mathbb{K}} a) + 0_{\mathbb{K}} a$. But $-(0_{\mathbb{K}} a) + 0_{\mathbb{K}} a = 0_{\mathbb{K}}$ (by the definition of $-(0_{\mathbb{K}} a)$), and associativity of addition shows that $-(0_{\mathbb{K}} a) + (0_{\mathbb{K}} a + 0_{\mathbb{K}} a) = \underbrace{(-(0_{\mathbb{K}} a) + 0_{\mathbb{K}} a)}_{= 0_{\mathbb{K}}} + 0_{\mathbb{K}} a = 0_{\mathbb{K}} + 0_{\mathbb{K}} a = 0_{\mathbb{K}} a$ (by neutrality of $0_{\mathbb{K}}$), so that $0_{\mathbb{K}} a = -(0_{\mathbb{K}} a) + (0_{\mathbb{K}} a + 0_{\mathbb{K}} a) = -(0_{\mathbb{K}} a) + 0_{\mathbb{K}} a = 0_{\mathbb{K}}$. Thus, $0_{\mathbb{K}} a = 0_{\mathbb{K}}$ is proven. Similarly one can show $a 0_{\mathbb{K}} = 0_{\mathbb{K}}$. Therefore, annihilation follows from the other axioms.

when this can cause no confusion (and, unfortunately, often also when it can). They are not always identical with the actual integers 0 and 1.

The binary operations $+$ and $\cdot$ in Definition 6.2 are also usually not identical with the binary operations $+$ and $\cdot$ on the set of integers, and are denoted by $+_{\mathbb{K}}$ and $\cdot_{\mathbb{K}}$ when confusion can arise.

The set $\mathbb{K}$ is called the *underlying set* of the commutative ring $\mathbb{K}$. Let us again remind ourselves that the underlying set of a commutative ring $\mathbb{K}$ is just a part of the data of $\mathbb{K}$.

Here are some examples and non-examples of rings:[176]

- The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ (endowed with the usual addition, the usual multiplication, the usual 0 and the usual 1) are commutative rings. (Notice that existence of **multiplicative** inverses is not required[177]!)

- The set $\mathbb{N}$ of nonnegative integers (again endowed with the usual addition, the usual multiplication, the usual 0 and the usual 1) is **not** a commutative ring. It fails the existence of additive inverses. (Of course, negative numbers exist, but this does not count because they don't lie in $\mathbb{N}$.)

- We can define a commutative ring $\mathbb{Z}'$ as follows: We define a binary operation $\widetilde{\times}$ on $\mathbb{Z}$ (written infix) by
$$\left(a \widetilde{\times} b = -ab \qquad \text{for all } (a,b) \in \mathbb{Z} \times \mathbb{Z} \right).$$
Now, let $\mathbb{Z}'$ be the **set** $\mathbb{Z}$, endowed with the usual addition $+$ and the (unusual) multiplication $\widetilde{\times}$, with the zero $0_{\mathbb{Z}'} = 0$ and with the unity $1_{\mathbb{Z}'} = -1$. It is easy to check that $\mathbb{Z}'$ is a commutative ring[178]; it is an example of a commutative ring whose unity is clearly **not** equal to the integer 1 (which is why it is important to never omit the subscript $\mathbb{Z}'$ in $1_{\mathbb{Z}'}$ here).

[175] Some people say "unit" instead of "unity", but other people use the word "unit" for something different, which makes every use of this word a potential pitfall.

[176] The following list of examples is long, and some of these examples rely on knowledge that you might not have yet. As usual with examples, you need not understand them all. When I say that Laurent polynomial rings are examples of commutative rings, I do not assume that you know what Laurent polynomials are; I merely want to ensure that, **if** you have already encountered Laurent polynomials, then you get to know that they form a commutative ring.

[177] A *multiplicative inverse* of an element $a \in \mathbb{K}$ means an element $a' \in \mathbb{K}$ such that $aa' = a'a = 1_{\mathbb{K}}$. (This is analogous to an additive inverse, except that addition is replaced by multiplication, and $0_{\mathbb{K}}$ is replaced by $1_{\mathbb{K}}$.) In a commutative ring, every element is required to have an additive inverse (by the definition of a commutative ring), but not every element is guaranteed to have a multiplicative inverse. (For instance, 2 has no multiplicative inverse in $\mathbb{Z}$, and 0 has no multiplicative inverse in any of the rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.)

We shall study multiplicative inverses more thoroughly in Section 6.8 (where we will just call them "inverses").

[178] Notice that we have named this new commutative ring $\mathbb{Z}'$, not $\mathbb{Z}$ (despite having $\mathbb{Z}' = \mathbb{Z}$ as sets). The reason is that if we had named it $\mathbb{Z}$, then we could no longer speak of "the commutative ring $\mathbb{Z}$" without being ambiguous (we would have to specify every time whether we mean the usual multiplication or the unusual one).

That said, $\mathbb{Z}'$ is not a very interesting ring: It is essentially "a copy of $\mathbb{Z}$, except that every integer n has been renamed as $-n$". To formalize this intuition, we would need to introduce the notion of a *ring isomorphism*, which we don't want to do right here; but the idea is that the bijection

$$\varphi : \mathbb{Z} \to \mathbb{Z}', \qquad n \mapsto -n$$

satisfies

$$\begin{aligned} \varphi(a+b) &= \varphi(a) + \varphi(b) && \text{for all } (a,b) \in \mathbb{Z} \times \mathbb{Z}; \\ \varphi(a \cdot b) &= \varphi(a) \widetilde{\times} \varphi(b) && \text{for all } (a,b) \in \mathbb{Z} \times \mathbb{Z}; \\ \varphi(0) &= 0_{\mathbb{Z}'}; \\ \varphi(1) &= 1_{\mathbb{Z}'}, \end{aligned}$$

and thus the ring $\mathbb{Z}'$ can be viewed as the ring $\mathbb{Z}$ with its elements "relabelled" using this bijection.

- The polynomial rings $\mathbb{Z}[x]$, $\mathbb{Q}[a,b]$, $\mathbb{C}[z_1, z_2, \ldots, z_n]$ are commutative rings. Laurent polynomial rings are also commutative rings. (Do not worry if you have not seen these rings yet.)

- The set of all functions $\mathbb{Q} \to \mathbb{Q}$ is a commutative ring, where addition and multiplication are defined pointwise (i.e., addition is defined by $(f+g)(x) = f(x) + g(x)$ for all $x \in \mathbb{Q}$, and multiplication is defined by $(fg)(x) = f(x) \cdot g(x)$ for all $x \in \mathbb{Q}$), where the zero is the "constant-0" function (sending every $x \in \mathbb{Q}$ to 0), and where the unity is the "constant-1" function (sending every $x \in \mathbb{Q}$ to 1). Of course, the same construction works if we consider functions $\mathbb{R} \to \mathbb{C}$, or functions $\mathbb{C} \to \mathbb{Q}$, or functions $\mathbb{N} \to \mathbb{Q}$, instead of functions $\mathbb{Q} \to \mathbb{Q}$. [179]

- The set $\mathbb{S}$ of all real numbers of the form $a + b\sqrt{5}$ with $a, b \in \mathbb{Q}$ (endowed with the usual notions of "addition" and "multiplication" defined on $\mathbb{R}$) is a commutative ring[180].

[179] But not if we consider functions $\mathbb{Q} \to \mathbb{N}$; such functions might fail the existence of additive inverses.

Generally, if X is any set and $\mathbb{K}$ is any commutative ring, then the set of all functions $X \to \mathbb{K}$ is a commutative ring, where addition and multiplication are defined pointwise, where the zero is the "constant-$0_{\mathbb{K}}$" function, and where the unity is the "constant-$1_{\mathbb{K}}$" function.

[180] To prove this, we argue as follows:

- Addition and multiplication are indeed two binary operations on $\mathbb{S}$. This is because the sum of two elements of $\mathbb{S}$ is an element of $\mathbb{S}$ (namely, $\left(a + b\sqrt{5}\right) + \left(c + d\sqrt{5}\right) = (a+c) + (b+d)\sqrt{5}$), and so is their product (namely, $\left(a + b\sqrt{5}\right) \cdot \left(c + d\sqrt{5}\right) = (ac + 5bd) + (bc + ad)\sqrt{5}$).
- All axioms of a commutative ring are satisfied for $\mathbb{S}$, except maybe the existence of additive inverses. This is simply because the addition and the multiplication in $\mathbb{S}$ are "inherited" from $\mathbb{R}$, and clearly all these axioms come with the inheritance.

- We could define a different ring structure on the set $\mathbb{S}$ (that is, a commutative ring which, as a set, is identical with $\mathbb{S}$, but has a different choice of operations) as follows: We define a binary operation $*$ on $\mathbb{S}$ by setting

  $$\left(a+b\sqrt{5}\right)*\left(c+d\sqrt{5}\right)=ac+bd\sqrt{5} \qquad \text{for all } (a,b)\in\mathbb{Q}\times\mathbb{Q} \text{ and } (c,d)\in\mathbb{Q}\times\mathbb{Q}.$$

  [181] Now, let $\mathbb{S}'$ be the set $\mathbb{S}$, endowed with the usual addition $+$ and the (unusual) multiplication $*$, with the zero $0_{\mathbb{S}'}=0$ and with the unity $1_{\mathbb{S}'}=1+\sqrt{5}$ (not the integer 1). It is easy to check that $\mathbb{S}'$ is a commutative ring[182]. The **sets** $\mathbb{S}$ and $\mathbb{S}'$ are identical, but the **commutative rings** $\mathbb{S}$ and $\mathbb{S}'$ are not[183]: For example, the ring $\mathbb{S}'$ has two nonzero elements whose product is 0 (namely, $1*\sqrt{5}=0$), whereas the ring $\mathbb{S}$ has no such things. This shows that not only do we have $\mathbb{S}'\neq\mathbb{S}$ as commutative rings, but there is also no way to regard $\mathbb{S}'$ as "a copy of $\mathbb{S}$ with its elements renamed" (in the same way as we have regarded $\mathbb{Z}'$ as "a copy of $\mathbb{Z}$ with its elements renamed"). This example should stress the point that a commutative ring $\mathbb{K}$ is not just a set; it is a set endowed with two operations ($+$ and $\cdot$) and two elements ($0_{\mathbb{K}}$ and $1_{\mathbb{K}}$), and these operations and elements are no less important than the set.

- The set $\mathbb{S}_3$ of all real numbers of the form $a+b\sqrt[3]{5}$ with $a,b\in\mathbb{Q}$ (endowed with the usual addition, the usual multiplication, the usual 0 and the usual 1) is **not** a commutative ring. Indeed, multiplication is not a binary operation on this set $\mathbb{S}_3$: It does not always send two elements of $\mathbb{S}_3$ to an element of $\mathbb{S}_3$. For instance, $\left(1+1\sqrt[3]{5}\right)\left(1+1\sqrt[3]{5}\right)=1+2\sqrt[3]{5}+\left(\sqrt[3]{5}\right)^2$ is not in $\mathbb{S}_3$.

- The set of all 2×2-matrices over $\mathbb{Q}$ is **not** a commutative ring, because commutativity of multiplication does not hold for this set. (In general, $AB\neq BA$ for matrices.)

- If you like the empty set, you will enjoy the *zero ring*. This is the commutative ring which is defined as the one-element set $\{0\}$, with zero and unity both being 0 (nobody said that they have to be distinct!), with addition given by $0+0=0$ and with multiplication given by $0\cdot 0=0$. Of course, it is not an empty set[184], but it plays a similar role in the world of commutative rings as

---

  – Existence of additive inverses also holds in $\mathbb{S}$, because the additive inverse of $a+b\sqrt{5}$ is $(-a)+(-b)\sqrt{5}$.

[181] This is well-defined, because every element of $\mathbb{S}$ can be written in the form $a+b\sqrt{5}$ for a **unique** pair $(a,b)\in\mathbb{Q}\times\mathbb{Q}$. This is a consequence of the irrationality of $\sqrt{5}$.

[182] Again, we do not call it $\mathbb{S}$, in order to be able to distinguish between different ring structures.

[183] Keep in mind that, due to our "pars pro toto" notation, "commutative ring $\mathbb{S}$" means more than "set $\mathbb{S}$".

[184] A commutative ring cannot be empty, as it contains at least one element (namely, 0).

the empty set does in the world of sets: It carries no information itself, but things would break if it were to be excluded[185].

Notice that the zero and the unity of the zero ring are identical, i.e., we have $0_{\mathbb{K}} = 1_{\mathbb{K}}$. This shows why it is dangerous to omit the subscripts and just denote the zero and the unity by 0 and 1; in fact, you don't want to rewrite the equality $0_{\mathbb{K}} = 1_{\mathbb{K}}$ as "$0 = 1$"! (Most algebraists make a compromise between wanting to omit the subscripts and having to clarify what 0 and 1 mean: They say that "$0 = 1$ in $\mathbb{K}$" to mean "$0_{\mathbb{K}} = 1_{\mathbb{K}}$".)

Generally, a *trivial ring* is defined to be a commutative ring containing only one element (which then necessarily is both the zero and the unity of this ring). The addition and the multiplication of a trivial ring are uniquely determined (since there is only one possible value that a sum or a product could take). Every trivial ring can be viewed as the zero ring with its element 0 relabelled.[186]

- In set theory, the *symmetric difference* of two sets A and B is defined to be the set $(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$. This symmetric difference is denoted by $A \triangle B$. Now, let S be any set. Let $\mathcal{P}(S)$ denote the powerset of S (that is, the set of all subsets of S). It is easy to check that the following ten properties hold:

$$A \triangle B = B \triangle A \qquad \text{for any sets } A \text{ and } B;$$
$$A \cap B = B \cap A \qquad \text{for any sets } A \text{ and } B;$$
$$(A \triangle B) \triangle C = A \triangle (B \triangle C) \qquad \text{for any sets } A, B \text{ and } C;$$
$$(A \cap B) \cap C = A \cap (B \cap C) \qquad \text{for any sets } A, B \text{ and } C;$$
$$A \triangle \varnothing = \varnothing \triangle A = A \qquad \text{for any set } A;$$
$$A \triangle A = \varnothing \qquad \text{for any set } A;$$
$$A \cap S = S \cap A = A \qquad \text{for any subset } A \text{ of } S;$$
$$\varnothing \cap A = A \cap \varnothing = \varnothing \qquad \text{for any set } A;$$
$$A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C) \qquad \text{for any sets } A, B \text{ and } C;$$
$$(A \triangle B) \cap C = (A \cap C) \triangle (B \cap C) \qquad \text{for any sets } A, B \text{ and } C.$$

Therefore, $\mathcal{P}(S)$ becomes a commutative ring, where the addition is defined to be the operation $\triangle$, the multiplication is defined to be the operation $\cap$, the zero is defined to be the set $\varnothing$, and the unity is defined to be the set S. [187]

[185]Some authors **do** prohibit the zero ring from being a commutative ring (by requiring every commutative ring to satisfy $0 \neq 1$). I think most of them run into difficulties from this decision sooner or later.

[186]In more formal terms, the preceding statement would say that "every trivial ring is isomorphic to the zero ring".

[187]The ten properties listed above show that the axioms of a commutative ring are satisfied for $(\mathcal{P}(S), \triangle, \cap, \varnothing, S)$. In particular, the sixth property shows that every subset A of S has an additive inverse – namely, itself. Of course, it is unusual for an element of a commutative ring to be its own additive inverse, but in this example it happens all the time!

The commutative ring $\mathcal{P}(S)$ has the property that $a \cdot a = a$ for every $a \in \mathcal{P}(S)$. (This simply means that $A \cap A = A$ for every $A \subseteq S$.) Commutative rings that have this property are called *Boolean rings*. (Of course, $\mathcal{P}(S)$ is the eponymic example for a Boolean ring; but there are also others.)

- For every positive integer n, the residue classes of integers modulo n form a commutative ring, which is called $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Z}_n$ (depending on the author). This ring has n elements (often called "integers modulo n"). When n is a composite number (e.g., $n = 6$), this ring has the property that products of nonzero[188] elements can be zero (e.g., we have $2 \cdot 3 \equiv 0 \bmod 6$); this means that there is no way to define division by all nonzero elements in this ring (even if we are allowed to create fractions). Notice that $\mathbb{Z}/1\mathbb{Z}$ is a trivial ring.

  We notice that if n is a positive integer, and if $\mathbb{K}$ is the commutative ring $\mathbb{Z}/n\mathbb{Z}$, then $\underbrace{1_{\mathbb{K}} + 1_{\mathbb{K}} + \cdots + 1_{\mathbb{K}}}_{n \text{ times}} = 0_{\mathbb{K}}$ (because the left hand side of this equality is the residue class of n modulo n, while the right hand side is the residue class of 0 modulo n, and these two residue classes are clearly equal).

- Let us try to define "division by zero". So, we introduce a new symbol ∞, and we try to extend the addition on $\mathbb{Q}$ to the set $\mathbb{Q} \cup \{\infty\}$ by setting $a + \infty = \infty$ for all $a \in \mathbb{Q} \cup \{\infty\}$. We might also try to extend the multiplication in some way, and perhaps to add some more elements (such as another symbol $-\infty$ to serve as the product $(-1)\infty$). I claim that (whatever we do with the multiplication, and whatever new elements we add) we do not get a commutative ring. Indeed, assume the contrary. Thus, there exists a commutative ring $\mathbb{W}$ which contains $\mathbb{Q} \cup \{\infty\}$ as a subset, and which has $a + \infty = \infty$ for all $a \in \mathbb{Q}$. Thus, in $\mathbb{W}$, we have $1 + \infty = \infty = 0 + \infty$. Adding $(-1)\infty$ to both sides of this equality, we obtain $1 + \infty + (-1)\infty = 0 + \infty + (-1)\infty$, so that $1 = 0$ [189]; but this is absurd. Hence, we have found a contradiction. This is why "division by zero is impossible": One can define objects that behave like "infinity" (and they **are** useful), but they break various standard rules such as the axioms of a commutative ring. In contrast to this, adding a "number" i satisfying $i^2 = -1$ to the real numbers is harmless: The complex numbers $\mathbb{C}$ are still a commutative ring.

- Here is an "almost-ring" beloved to many combinatorialists: the *max-plus semiring* $\mathbb{T}$ (also called the *tropical semiring*[190]). We create a new symbol $-\infty$, and we set $\mathbb{T} = \mathbb{Z} \cup \{-\infty\}$ as sets, but we do **not** "inherit" the addition and the multiplication from $\mathbb{Z}$. Instead, we denote the "addition" and "multiplication" operations on $\mathbb{Z}$ by $+_{\mathbb{Z}}$ and $\cdot_{\mathbb{Z}}$, and we define two new "addition"

[188] An element a of a commutative ring $\mathbb{K}$ is said to be *nonzero* if $a \neq 0_{\mathbb{K}}$. (This is not the same as saying that a is not the integer 0, because the integer 0 might not be $0_{\mathbb{K}}$.)

[189] because $\infty + (-1)\infty = 1\infty + (-1)\infty = \underbrace{(1 + (-1))}_{=0}\infty = 0\infty = 0$

[190] Caution: Both of these names mean many other things as well.

and "multiplication" operations $+_{\mathbb{T}}$ and $\cdot_{\mathbb{T}}$ on $\mathbb{T}$ as follows:

$$a +_{\mathbb{T}} b = \max\{a, b\};$$
$$a \cdot_{\mathbb{T}} b = a +_{\mathbb{Z}} b.$$

(Here, we set $\max\{-\infty, n\} = \max\{n, -\infty\} = n$ and $(-\infty) +_{\mathbb{Z}} n = n +_{\mathbb{Z}} (-\infty) = -\infty$ for every $n \in \mathbb{T}$.)

It turns out that the set $\mathbb{T}$ endowed with the two operations $+_{\mathbb{T}}$ and $\cdot_{\mathbb{T}}$, the zero $0_{\mathbb{T}} = -\infty$ and the unity $1_{\mathbb{T}} = 0$ comes rather close to being a commutative ring. It satisfies all axioms of a commutative ring except for the existence of additive inverses. Such a structure is called a *semiring*. Other examples of semirings are $\mathbb{N}$ and a reasonably defined $\mathbb{N} \cup \{\infty\}$ (with $0\infty = 0$ and $a\infty = \infty$ for all $a > 0$).

If $\mathbb{K}$ is a commutative ring, then we can define a subtraction in $\mathbb{K}$, even though we have not required a subtraction operation as part of the definition of a commutative ring $\mathbb{K}$. Namely, the *subtraction* of a commutative ring $\mathbb{K}$ is the binary operation $-$ on $\mathbb{K}$ (again written infix) defined as follows: For every $a \in \mathbb{K}$ and $b \in \mathbb{K}$, set $a - b = a + b'$, where b' is the additive inverse of b. It is not hard to check that $a - b$ is the unique element c of $\mathbb{K}$ satisfying $a = b + c$; thus, subtraction is "the undoing of addition" just as in the classical situation of integers. Again, the notation $-$ for the subtraction of $\mathbb{K}$ is denoted by $-_{\mathbb{K}}$ whenever a confusion with the subtraction of integers could arise.

Whenever a is an element of a commutative ring $\mathbb{K}$, we write $-a$ for the additive inverse of a. This is the same as $0_{\mathbb{K}} - a$.

The intuition for commutative rings is essentially that all computations that can be performed with the operations $+$, $-$ and $\cdot$ on integers can be similarly made in any commutative ring. For instance, if $a_1, a_2, \ldots, a_n$ are n elements of a commutative ring, then the sum $a_1 + a_2 + \cdots + a_n$ is well-defined, and can be computed by adding the elements $a_1, a_2, \ldots, a_n$ to each other in any order[191]. More generally: If S is a finite set, if $\mathbb{K}$ is a commutative ring, and if $(a_s)_{s \in S}$ is a $\mathbb{K}$-valued S-family[192], then the sum $\sum\limits_{s \in S} a_s$ is defined in the same way as finite sums of numbers were defined in Section 1.4 (but with $\mathbb{A}$ replaced by $\mathbb{K}$, of course[193]); this definition is still legitimate[194], and these finite sums of elements of $\mathbb{K}$ satisfy the same properties as finite sums of numbers (see Section 1.4 for these properties). All this can be proven in the same way as it was proven for numbers (in Section 2.14 and Section

[191] For instance, we can compute the sum $a + b + c + d$ of four elements a, b, c, d in many ways: For example, we can first add a and b, then add c and d, and finally add the two results; alternatively, we can first add a and b, then add d to the result, then add c to the result. In a commutative ring, all such ways lead to the same result.

[192] See Definition 2.107 for the definition of this notion.

[193] and, consequently, 0 replaced by $0_{\mathbb{K}}$

[194] i.e., the result does not depend on the choice of t in (1)

1.4). The same holds for finite products. Furthermore, if n is an integer and a is an element of a commutative ring $\mathbb{K}$, then we define an element na of $\mathbb{K}$ by

$$na = \begin{cases} \underbrace{a + a + \cdots + a}_{n \text{ addends}}, & \text{if } n \geq 0; \\ -\left(\underbrace{a + a + \cdots + a}_{-n \text{ addends}}\right), & \text{if } n < 0 \end{cases}.$$

[195]

If n is a nonnegative integer and a is an element of a commutative ring $\mathbb{K}$, then a^n is a well-defined element of $\mathbb{K}$ (namely, $a^n = \underbrace{a \cdot a \cdot \cdots \cdot a}_{n \text{ factors}}$). In particular, applying this definition to $n = 0$, we obtain

$$a^0 = \underbrace{a \cdot a \cdot \cdots \cdot a}_{0 \text{ factors}} = (\text{empty product}) = 1 \qquad \text{for each } a \in \mathbb{K}.$$

The following identities hold:

$$(n+m)\,a = na + ma \qquad \text{for } a \in \mathbb{K} \text{ and } n, m \in \mathbb{Z}; \tag{325}$$

$$n\,(a+b) = na + nb \qquad \text{for } a, b \in \mathbb{K} \text{ and } n \in \mathbb{Z}; \tag{326}$$

$$-\,(a+b) = (-a) + (-b) \qquad \text{for } a, b \in \mathbb{K}; \tag{327}$$

$$1a = a \qquad \text{for } a \in \mathbb{K}; \tag{328}$$

$$0a = 0_{\mathbb{K}} \qquad \text{for } a \in \mathbb{K}$$
$$(\text{here, the ``0'' on the left hand side means the integer } 0);$$

$$(-1)\,a = -a \qquad \text{for } a \in \mathbb{K}; \tag{329}$$

$$-\,(-a) = a \qquad \text{for } a \in \mathbb{K}; \tag{330}$$

$$-\,(ab) = (-a)\,b = a\,(-b) \qquad \text{for } a, b \in \mathbb{K}; \tag{331}$$

$$-\,(na) = (-n)\,a = n\,(-a) \qquad \text{for } a \in \mathbb{K} \text{ and } n \in \mathbb{Z}; \tag{332}$$

$$n\,(ab) = (na)\,b = a\,(nb) \qquad \text{for } a, b \in \mathbb{K} \text{ and } n \in \mathbb{Z}; \tag{333}$$

$$(nm)\,a = n\,(ma) \qquad \text{for } a \in \mathbb{K} \text{ and } n, m \in \mathbb{Z}; \tag{334}$$

$$n0_{\mathbb{K}} = 0_{\mathbb{K}} \qquad \text{for } n \in \mathbb{Z};$$

$$1^n = 1 \qquad \text{for } n \in \mathbb{N};$$

$$0^n = \begin{cases} 0, & \text{if } n > 0; \\ 1, & \text{if } n = 0 \end{cases} \qquad \text{for } n \in \mathbb{N}; \tag{335}$$

$$a^{n+m} = a^n a^m \qquad \text{for } a \in \mathbb{K} \text{ and } n, m \in \mathbb{N}; \tag{336}$$

$$a^{nm} = (a^n)^m \qquad \text{for } a \in \mathbb{K} \text{ and } n, m \in \mathbb{N};$$

$$(ab)^n = a^n b^n \qquad \text{for } a, b \in \mathbb{K} \text{ and } n \in \mathbb{N}. \tag{337}$$

[195] Notice that this definition of na is **not** a particular case of the product of two elements of $\mathbb{K}$, because n is not an element of $\mathbb{K}$.

Here, we are using the standard notations $+$, $\cdot$, 0 and 1 for the addition, the multiplication, the zero and the unity of $\mathbb{K}$, because confusion (e.g., confusion of the 0 with the integer 0) is rather unlikely.[196] We shall keep doing so in the following, apart from situations where confusion can realistically occur.[197]

The identities listed above are not hard to prove. Indeed, they are generalizations of well-known identities holding for rational numbers; and some of them (for example, (336) and (337)) can be proved in exactly the same way as those identities for rational numbers.[198]

If a and b are two elements of a commutative ring $\mathbb{K}$, then the expression "$-ab$" appears ambiguous, since it can be interpreted either as "$-(ab)$" or as "$(-a)b$". But (331) shows that these two interpretations yield the same result; thus, we can write this expression "$-ab$" without fearing ambiguity. Similarly, if $n \in \mathbb{Z}$ and $a, b \in \mathbb{K}$, then the expression "nab" is unambiguous, because (333) shows that the two possible ways to interpret it (namely, as "$n(ab)$" and as "$(na)b$") yield the same result. Similarly, if $n, m \in \mathbb{Z}$ and $a \in \mathbb{K}$, then the expression "nma" is unambiguous, because of (334).

Furthermore, finite sums such as $\sum\limits_{s \in S} a_s$ (where S is a finite set, and $a_s \in \mathbb{K}$ for every $s \in S$), and finite products such as $\prod\limits_{s \in S} a_s$ (where S is a finite set, and $a_s \in \mathbb{K}$ for every $s \in S$) are defined whenever $\mathbb{K}$ is a commutative ring. Again, the definition is the same as for numbers, and these sums and products behave as they do for numbers.[199] For example, Exercise 5.13 still holds if we replace "$\mathbb{C}$" by "$\mathbb{K}$" in it (and the same solution proves it) whenever $\mathbb{K}$ is a commutative ring. From the fact that finite sums and finite products of elements of $\mathbb{K}$ are well-defined, we can also conclude that expressions such as "$a_1 + a_2 + \cdots + a_k$" and "$a_1 a_2 \cdots a_k$" (where $a_1, a_2, \ldots, a_k$ are finitely many elements of $\mathbb{K}$) are well-defined.

---

[196] For instance, in the statement "$-(a+b) = (-a) + (-b)$ for $a, b \in \mathbb{K}$", it is clear that the $+$ can only stand for the addition of $\mathbb{K}$ and not (say) for the addition of integers (since a, b, $-a$ and $-b$ are elements of $\mathbb{K}$, not (generally) integers). The only statement whose meaning is ambiguous is "$0^n = \begin{cases} 0, & \text{if } n > 0; \\ 1, & \text{if } n = 0 \end{cases}$ for $n \in \mathbb{N}$". In this statement, the "0" in "$n > 0$" and the "0" in "$n = 0$" clearly mean the integer 0 (since they are being compared with the integer n), but the other two appearances of "0" and the "1" are ambiguous. I hope that the context makes it clear enough that they mean the zero and the unity of $\mathbb{K}$ (and not the integers 0 and 1), because otherwise this equality would not be a statement about $\mathbb{K}$ at all.

[197] Notice that the equalities (333) and (334) are **not** particular cases of the associativity of multiplication which we required to hold for $\mathbb{K}$. Indeed, the latter associativity says that $a(bc) = (ab)c$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$. But in (333) and (334), the n is an integer, not an element of $\mathbb{K}$.

[198] For example, it is well-known that

$$(ab)^n = a^n b^n \qquad \text{for any } a, b \in \mathbb{Q} \text{ and } n \in \mathbb{N}.$$

This can be easily proven by induction on n, using the commutativity and associativity rules for multiplication of rational numbers and the fact that $1 \cdot 1 = 1$. The same argument can be used to prove (337). The only change required is replacing every appearance of "$\mathbb{Q}$" by "$\mathbb{K}$".

[199] Of course, empty sums of elements of $\mathbb{K}$ are defined to equal $0_{\mathbb{K}}$, and empty products of elements of $\mathbb{K}$ are defined to equal $1_{\mathbb{K}}$.

Various identities that hold for numbers also hold for elements of arbitrary commutative rings. For example, an analogue of the binomial formula (Proposition 3.21) holds: If $\mathbb{K}$ is a commutative ring, then

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \qquad \text{for } a, b \in \mathbb{K} \text{ and } n \in \mathbb{N}. \tag{338}$$

(We can obtain a proof of (338) by re-reading the solution to Exercise 3.6, while replacing every "x" by an "a" and replacing every "y" by a "b". Another proof of (338) is given in the solution to Exercise 6.1 **(b)**.)

Remark 6.4. The notion of a "commutative ring" is not fully standardized; there exist several competing definitions:

For some people, a "commutative ring" is **not** endowed with an element 1 (although it **can** have such an element), and, consequently, does not have to satisfy the unitality axiom. According to their definition, for example, the set

$$\{\ldots, -4, -2, 0, 2, 4, \ldots\} = \{2n \mid n \in \mathbb{Z}\} = (\text{the set of all even integers})$$

is a commutative ring (with the usual addition and multiplication). (In contrast, our definition of a "commutative ring" does not accept this set as a commutative ring, because it does not contain any element which would fill the role of 1.) These people tend to use the notation "commutative ring with unity" (or "commutative ring with 1") to mean a commutative ring which is endowed with a 1 and satisfies the unitality axiom (i.e., what we call a "commutative ring").

On the other hand, there are authors who use the word "ring" for what we call "commutative ring". These are mostly the authors who work with commutative rings all the time and find the name "commutative ring" too long.

When you are reading about rings, it is important to know which meaning of "ring" the author is subscribing to. (Often this can be inferred from the examples given.)

Exercise 6.1. Let $\mathbb{K}$ be a commutative ring. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

(a) Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n further elements of $\mathbb{K}$. Prove that

$$\prod_{i=1}^{n} (a_i + b_i) = \sum_{I \subseteq [n]} \left(\prod_{i \in I} a_i \right) \left(\prod_{i \in [n] \setminus I} b_i \right).$$

(Here, as usual, the summation sign $\sum_{I \subseteq [n]}$ means $\sum_{I \in \mathcal{P}([n])}$, where $\mathcal{P}([n])$ denotes the powerset of $[n]$.)

(b) Use Exercise 6.1 to give a new proof of (338).

Exercise 6.2. For each $m \in \mathbb{N}$ and $(k_1, k_2, \ldots, k_m) \in \mathbb{N}^m$, let us define a positive integer $\mathbf{m}(k_1, k_2, \ldots, k_m)$ by $\mathbf{m}(k_1, k_2, \ldots, k_m) = \dfrac{(k_1 + k_2 + \cdots + k_m)!}{k_1! k_2! \cdots k_m!}$. (This is indeed a positive integer, because Exercise 3.1 says so.)

Let $\mathbb{K}$ be a commutative ring. Let $m \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_m$ be m elements of $\mathbb{K}$. Let $n \in \mathbb{N}$. Prove that

$$(a_1 + a_2 + \cdots + a_m)^n = \sum_{\substack{(k_1, k_2, \ldots, k_m) \in \mathbb{N}^m; \\ k_1 + k_2 + \cdots + k_m = n}} \mathbf{m}(k_1, k_2, \ldots, k_m) \prod_{i=1}^{m} a_i^{k_i}.$$

(This is called the *multinomial formula*.)

6.2. Matrices

We have briefly defined determinants in Definition 5.16, but we haven't done much with them. This will be amended now. But let us first recall the definitions of basic notions in matrix algebra.

In the following, we fix a commutative ring $\mathbb{K}$. The elements of $\mathbb{K}$ will be called *scalars* (to distinguish them from *vectors* and *matrices*, which we will soon discuss, and which are structures containing several elements of $\mathbb{K}$).

If you feel uncomfortable with commutative rings, you are free to think that $\mathbb{K} = \mathbb{Q}$ or $\mathbb{K} = \mathbb{C}$ in the following; but everything I am doing works for any commutative ring unless stated otherwise.

Given two nonnegative integers n and m, an *$n \times m$-matrix* (or, more precisely, *$n \times m$-matrix over* $\mathbb{K}$) means a rectangular table with n rows and m columns whose entries are elements of $\mathbb{K}$. [200] For instance, when $\mathbb{K} = \mathbb{Q}$, the table $\begin{pmatrix} 1 & -2/5 & 4 \\ 1/3 & -1/2 & 0 \end{pmatrix}$ is a 2×3-matrix. A *matrix* simply means an $n \times m$-matrix for some $n \in \mathbb{N}$ and $m \in \mathbb{N}$. These n and m are said to be the *dimensions* of the matrix.

If A is an $n \times m$-matrix, and if $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$, then the *$(i, j)$-th entry of A* means the entry of A in row i and column j. For instance, the $(1, 2)$-th entry of the matrix $\begin{pmatrix} 1 & -2/5 & 4 \\ 1/3 & -1/2 & 0 \end{pmatrix}$ is $-2/5$.

If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, and if we are given an element $a_{i,j} \in \mathbb{K}$ for every $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$, then we use the notation $(a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m}$ for the $n \times$

[200] Formally speaking, this means that an $n \times m$-matrix is a map from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to $\mathbb{K}$. We represent such a map as a rectangular table by writing the image of $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ into the cell in the i-th row and the j-th column.

Thus, the notion of an $n \times m$-matrix is closely akin to what we called an "$n \times m$-table of elements of $\mathbb{K}$" in Definition 2.110. The main difference between these two notions is that an $n \times m$-matrix "knows" $\mathbb{K}$, whereas an $n \times m$-table does not (i.e., two $n \times m$-matrices that have the same entries in the same positions but are defined using different commutative rings $\mathbb{K}$ are considered different, but two such $n \times m$-tables are considered identical).

m-matrix whose (i,j)-th entry is $a_{i,j}$ for all $(i,j) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,m\}$. Thus,

$$\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{pmatrix}.$$

The letters i and j are not set in stone; they are bound variables like the k in "$\sum\limits_{k=1}^{n} k$". Thus, you are free to write $\left(a_{x,y}\right)_{1\leq x\leq n,\ 1\leq y\leq m}$ or $\left(a_{j,i}\right)_{1\leq j\leq n,\ 1\leq i\leq m}$ instead of $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ (and we will use this freedom eventually).[201]

Matrices can be added if they share the same dimensions: If n and m are two nonnegative integers, and if $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ and $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ are two $n \times m$-matrices, then $A + B$ means the $n \times m$-matrix $\left(a_{i,j} + b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$. Thus, matrices are added "entry by entry"; for example, $\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} + \begin{pmatrix} a' & b' & c' \\ d' & e' & f' \end{pmatrix} = \begin{pmatrix} a+a' & b+b' & c+c' \\ d+d' & e+e' & f+f' \end{pmatrix}$. Similarly, subtraction is defined: If $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ and $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$, then $A - B = \left(a_{i,j} - b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$.

Similarly, one can define the product of a scalar $\lambda \in \mathbb{K}$ with a matrix A: If $\lambda \in \mathbb{K}$ is a scalar, and if $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ is an $n \times m$-matrix, then λA means the $n \times m$-matrix $\left(\lambda a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$.

Defining the product of two matrices is trickier. Matrices are **not** multiplied "entry by entry"; this would not be a very interesting definition. Instead, their product is defined as follows: If n, m and ℓ are three nonnegative integers, then the product AB of an $n \times m$-matrix $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ with an $m \times \ell$-matrix $B = \left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq \ell}$ means the $n \times \ell$-matrix

$$\left(\sum_{k=1}^{m} a_{i,k} b_{k,j}\right)_{1\leq i\leq n,\ 1\leq j\leq \ell}.$$

This definition looks somewhat counterintuitive, so let me comment on it. First of all, for AB to be defined, A and B are **not** required to have the same dimensions; instead, A must have as many columns as B has rows. The resulting matrix AB then has as many rows as A and as many columns as B. Every entry of AB is a sum of products of an entry of A with an entry of B (not a single such product).

[201] Many authors love to abbreviate "$a_{i,j}$" by "a_{ij}" (hoping that the reader will not mistake the subscript "ij" for a product or (in the case where i and j are single-digit numbers) for a two-digit number). The only advantage of this abbreviation that I am aware of is that it saves you a comma; I do not understand why it is so popular. But you should be aware of it in case you are reading other texts.

More precisely, the (i,j)-th entry of AB is a sum of products of an entry in the i-th row of A with the respective entry in the j-th column of B. For example,

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} a' & d' & g' \\ b' & e' & h' \\ c' & f' & i' \end{pmatrix} = \begin{pmatrix} aa'+bb'+cc' & ad'+be'+cf' & ag'+bh'+ci' \\ da'+eb'+fc' & dd'+ee'+ff' & dg'+eh'+fi' \end{pmatrix}.$$

The multiplication of matrices is not commutative! It is easy to find examples of two matrices A and B for which the products AB and BA are distinct, or one of them is well-defined but the other is not[202].

For given $n \in \mathbb{N}$ and $m \in \mathbb{N}$, we define the $n \times m$ *zero matrix* to be the $n \times m$-matrix whose all entries are 0 (that is, the $n \times m$-matrix $(0)_{1\leq i\leq n,\ 1\leq j\leq m}$). We denote this matrix by $0_{n\times m}$. If A is any $n \times m$-matrix, then the $n \times m$-matrix $-A$ is defined to be $0_{n\times m} - A$.

A sum $\sum\limits_{i\in I} A_i$ of finitely many matrices A_i is defined in the same way as a sum of numbers or of elements of a commutative ring[203]. However, a product $\prod\limits_{i\in I} A_i$ of finitely many matrices A_i (in general) cannot be defined, because the result would depend on the order of multiplication.

For every $n \in \mathbb{N}$, we let I_n denote the $n \times n$-matrix $\left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, where $\delta_{i,j}$ is defined to be $\begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$. [204] This matrix I_n looks as follows:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

It has the property that $I_n B = B$ for every $m \in \mathbb{N}$ and every $n \times m$-matrix B; also, $AI_n = A$ for every $k \in \mathbb{N}$ and every $k \times n$-matrix A. (Proving this is a good way to check that you understand how matrices are multiplied.[205]) The matrix I_n is called the $n \times n$ *identity matrix*. (Some call it E_n or just I, when the value of n is clear from the context.)

Matrix multiplication is associative: If $n, m, k, \ell \in \mathbb{N}$, and if A is an $n \times m$-matrix, B is an $m \times k$-matrix, and C is a $k \times \ell$-matrix, then $A(BC) = (AB)C$. The proof

[202] This happens if A has as many columns as B has rows, but B does not have as many columns as A has rows.

[203] with the caveat that an empty sum of $n \times m$-matrices is not the number 0, but the $n \times m$-matrix $0_{n,m}$

[204] Here, 0 and 1 mean the zero and the unity of $\mathbb{K}$ (which may and may not be the integers 0 and 1).

[205] See [Grinbe16b, §2.12] for a detailed proof of the equality $AI_n = A$. (Interpret the word "number" in [Grinbe16b, §2.12] as "element of $\mathbb{K}$".) The proof of $I_n B = B$ is rather similar.

of this is straightforward using our definition of products of matrices[206]. This associativity allows us to write products like ABC without parentheses. By induction, we can see that longer products such as $A_1 A_2 \cdots A_k$ for arbitrary $k \in \mathbb{N}$ can also be bracketed at will, because all bracketings lead to the same result (e.g., for four matrices A, B, C and D, we have $A(B(CD)) = A((BC)D) = (AB)(CD) = (A(BC))D = ((AB)C)D$, provided that the dimensions of the matrices are appropriate to make sense of the products). We define an empty product of $n \times n$-matrices to be the $n \times n$ identity matrix I_n.

For every $n \times n$-matrix A and every $k \in \mathbb{N}$, we can thus define an $n \times n$-matrix A^k by $A^k = \underbrace{AA \cdots A}_{k \text{ factors}}$. In particular, $A^0 = I_n$ (since we defined an empty product of $n \times n$-matrices to be I_n).

Further properties of matrix multiplication are easy to state and to prove:

- For every $n \in \mathbb{N}$, $m \in \mathbb{N}$, $k \in \mathbb{N}$ and $\lambda \in \mathbb{K}$, every $n \times m$-matrix A and every $m \times k$-matrix B, we have $\lambda(AB) = (\lambda A)B = A(\lambda B)$. (This allows us to write λAB for each of the matrices $\lambda(AB)$, $(\lambda A)B$ and $A(\lambda B)$.)

- For every $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $k \in \mathbb{N}$, every two $n \times m$-matrices A and B, and every $m \times k$-matrix C, we have $(A+B)C = AC + BC$.

- For every $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $k \in \mathbb{N}$, every $n \times m$-matrix A, and every two $m \times k$-matrices B and C, we have $A(B+C) = AB + AC$.

- For every $n \in \mathbb{N}$, $m \in \mathbb{N}$, $\lambda \in \mathbb{K}$ and $\mu \in \mathbb{K}$, and every $n \times m$-matrix A, we have $\lambda(\mu A) = (\lambda\mu)A$. (This allows us to write $\lambda\mu A$ for both $\lambda(\mu A)$ and $(\lambda\mu)A$.)

For given $n \in \mathbb{N}$ and $m \in \mathbb{N}$, we let $\mathbb{K}^{n \times m}$ denote the set of all $n \times m$-matrices. (This is one of the two standard notations for this set; the other is $\mathrm{M}_{n,m}(\mathbb{K})$.)

A *square matrix* is a matrix which has as many rows as it has columns; in other words, a square matrix is an $n \times n$-matrix for some $n \in \mathbb{N}$. If $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ is a square matrix, then the n-tuple $(a_{1,1}, a_{2,2}, \ldots, a_{n,n})$ is called the *diagonal* of A. (Some authors abbreviate $(a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ by $(a_{i,j})_{1 \leq i,j \leq n}$; this notation has some mild potential for confusion, though[207].) The entries of the diagonal of A are called the *diagonal entries* of A. (Some authors like to say "main diagonal" instead of "diagonal".)

---

[206]Check that $A(BC)$ and $(AB)C$ both are equal to the matrix $\left(\sum_{u=1}^{m} \sum_{v=1}^{k} a_{i,u} b_{u,v} c_{v,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq \ell}$. For details of this proof, see [Grinbe16b, §2.9]. (Interpret the word "number" in [Grinbe16b, §2.9] as "element of $\mathbb{K}$".)

[207]The comma between "i" and "j" in "$1 \leq i, j \leq n$" can be understood either to separate i from j, or to separate the inequality $1 \leq i$ from the inequality $j \leq n$. I remember seeing this ambiguity causing a real misunderstanding.

For a given $n \in \mathbb{N}$, the product of two $n \times n$-matrices is always well-defined, and is an $n \times n$-matrix again. The set $\mathbb{K}^{n \times n}$ satisfies all the axioms of a commutative ring except for commutativity of multiplication. This makes it into what is commonly called a *noncommutative ring*[208]. We shall study noncommutative rings later (in Section 6.17).

6.3. Determinants

Square matrices have determinants. Let us recall how determinants are defined:

Definition 6.5. Let $n \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. The *determinant* $\det A$ of A is defined as

$$\sum_{\sigma \in S_n} (-1)^{\sigma} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}. \tag{339}$$

In other words,

$$\det A = \sum_{\sigma \in S_n} (-1)^{\sigma} \underbrace{a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}}_{= \prod_{i=1}^{n} a_{i,\sigma(i)}} \tag{340}$$

$$= \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}. \tag{341}$$

For example, the determinant of a 1×1-matrix $\left(a_{1,1} \right)$ is

$$\det \left(a_{1,1} \right) = \sum_{\sigma \in S_1} (-1)^{\sigma} a_{1,\sigma(1)} = \underbrace{(-1)^{\mathrm{id}}}_{=1} \underbrace{a_{1,\mathrm{id}(1)}}_{=a_{1,1}} \qquad \text{(since the only permutation } \sigma \in S_1 \text{ is id)}$$
$$= a_{1,1}. \tag{342}$$

[208] A *noncommutative ring* is defined in the same way as we defined a commutative ring, except for the fact that commutativity of multiplication is removed from the list of axioms. (The words "noncommutative ring" do not imply that commutativity of multiplication must be false for this ring; they merely say that commutativity of multiplication is **not required** to hold for it. For example, the noncommutative ring $\mathbb{K}^{n \times n}$ is actually commutative when $n \leq 1$ or when $\mathbb{K}$ is a trivial ring.)

Instead of saying "noncommutative ring", many algebraists just say "ring". We shall, however, keep the word "noncommutative" in order to avoid confusion.

The determinant of a 2×2-matrix $\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$ is

$$\begin{aligned} \det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} &= \sum_{\sigma \in S_2} (-1)^{\sigma} a_{1,\sigma(1)} a_{2,\sigma(2)} \\ &= \underbrace{(-1)^{\operatorname{id}}}_{=1} \underbrace{a_{1,\operatorname{id}(1)}}_{=a_{1,1}} \underbrace{a_{2,\operatorname{id}(2)}}_{=a_{2,2}} + \underbrace{(-1)^{s_1}}_{=-1} \underbrace{a_{1,s_1(1)}}_{=a_{1,2}} \underbrace{a_{2,s_1(2)}}_{=a_{2,1}} \\ &\qquad \left(\text{since the only permutations } \sigma \in S_2 \text{ are } \operatorname{id} \text{ and } s_1\right) \\ &= a_{1,1} a_{2,2} - a_{1,2} a_{2,1}. \end{aligned}$$

Similarly, for a 3×3-matrix, the formula is

$$\begin{aligned} \det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} &= a_{1,1} a_{2,2} a_{3,3} + a_{1,2} a_{2,3} a_{3,1} + a_{1,3} a_{2,1} a_{3,2} \\ &\qquad - a_{1,1} a_{2,3} a_{3,2} - a_{1,2} a_{2,1} a_{3,3} - a_{1,3} a_{2,2} a_{3,1}. \end{aligned} \tag{343}$$

Also, the determinant of the 0×0-matrix is 1 [209]. (This might sound like hair-splitting, but being able to work with 0×0-matrices simplifies some proofs by induction, because it allows one to take $n = 0$ as an induction base.)

The equality (341) (or, equivalently, (340)) is known as the *Leibniz formula*. Out of several known ways to define the determinant, it is probably the most direct. In practice, however, computing a determinant using (341) quickly becomes impractical when n is high (since the sum has $n!$ terms). In most situations that occur both in mathematics and in applications, determinants can be computed in various simpler ways.

Some authors write $|A|$ instead of $\det A$ for the determinant of a square matrix A. I do not like this notation, as it clashes (in the case of 1×1-matrices) with the notation $|a|$ for the absolute value of a real number a.

Here is a first example of a determinant which ends up very simple:

---

[209] In more details:

There is only one 0×0-matrix; it has no rows and no columns and no entries. According to (341), its determinant is

$$\begin{aligned} \sum_{\sigma \in S_0} (-1)^{\sigma} \underbrace{\prod_{i=1}^{0} a_{i,\sigma(i)}}_{=(\text{empty product})=1} &= \sum_{\sigma \in S_0} (-1)^{\sigma} = (-1)^{\operatorname{id}} \qquad \left(\text{since the only } \sigma \in S_0 \text{ is } \operatorname{id}\right) \\ &= 1. \end{aligned}$$

Example 6.6. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$, and let $y_1, y_2, \ldots, y_n$ be n further elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i y_j\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. What is $\det A$?

For $n = 0$, we have $\det A = 1$ (since the 0×0-matrix has determinant 1).

For $n = 1$, we have $A = \left(\begin{array}{c} x_1 y_1 \end{array}\right)$ and thus $\det A = x_1 y_1$.

For $n = 2$, we have $A = \left(\begin{array}{cc} x_1 y_1 & x_1 y_2 \\ x_2 y_1 & x_2 y_2 \end{array}\right)$ and thus $\det A = (x_1 y_1)(x_2 y_2) - (x_1 y_2)(x_2 y_1) = 0$.

What do you expect for greater values of n ? The pattern might not be clear at this point yet, but if you compute further examples, you will realize that $\det A = 0$ also holds for $n = 3$, for $n = 4$, for $n = 5$... This suggests that $\det A = 0$ for every $n \geq 2$. How to prove this?

Let $n \geq 2$. Then, (340) (applied to $a_{i,j} = x_i y_j$) yields

$$\begin{aligned}
\det A &= \sum_{\sigma \in S_n} (-1)^{\sigma} \underbrace{\left(x_1 y_{\sigma(1)}\right)\left(x_2 y_{\sigma(2)}\right) \cdots \left(x_n y_{\sigma(n)}\right)}_{=(x_1 x_2 \cdots x_n)\left(y_{\sigma(1)} y_{\sigma(2)} \cdots y_{\sigma(n)}\right)} \\
&= \sum_{\sigma \in S_n} (-1)^{\sigma} (x_1 x_2 \cdots x_n) \underbrace{\left(y_{\sigma(1)} y_{\sigma(2)} \cdots y_{\sigma(n)}\right)}_{\substack{= y_1 y_2 \cdots y_n \\ \text{(since } \sigma \text{ is a permutation)}}} \\
&= \sum_{\sigma \in S_n} (-1)^{\sigma} (x_1 x_2 \cdots x_n)(y_1 y_2 \cdots y_n) \\
&= \left(\sum_{\sigma \in S_n} (-1)^{\sigma}\right)(x_1 x_2 \cdots x_n)(y_1 y_2 \cdots y_n). \qquad (344)
\end{aligned}$$

Now, every $\sigma \in S_n$ is either even or odd (but not both), and thus we have

$$\begin{aligned}
&\sum_{\sigma \in S_n} (-1)^{\sigma} \\
&= \sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is even}}} \underbrace{(-1)^{\sigma}}_{\substack{=1 \\ \text{(since } \sigma \text{ is even)}}} + \sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is odd}}} \underbrace{(-1)^{\sigma}}_{\substack{=-1 \\ \text{(since } \sigma \text{ is odd)}}} \\
&= \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is even}}} 1}_{=(\text{the number of even permutations } \sigma \in S_n) \cdot 1} + \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is odd}}} (-1)}_{=(\text{the number of odd permutations } \sigma \in S_n) \cdot (-1)} \\
&= \underbrace{(\text{the number of even permutations } \sigma \in S_n)}_{\substack{= n!/2 \\ \text{(by Exercise 5.4)}}} \cdot 1 \\
&\qquad + \underbrace{(\text{the number of odd permutations } \sigma \in S_n)}_{\substack{= n!/2 \\ \text{(by Exercise 5.4)}}} \cdot (-1) \\
&= (n!/2) \cdot 1 + (n!/2) \cdot (-1) = 0.
\end{aligned}$$

Hence, (344) becomes $\det A = \underbrace{\left(\sum_{\sigma \in S_n} (-1)^\sigma\right)}_{=0} (x_1 x_2 \cdots x_n)(y_1 y_2 \cdots y_n) = 0$, as we wanted to prove.

We will eventually learn a simpler way to prove this.

Example 6.7. Here is an example similar to Example 6.6, but subtler.

Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$, and let $y_1, y_2, \ldots, y_n$ be n further elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $(x_i + y_j)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. What is $\det A$?

For $n = 0$, we have $\det A = 1$ again.

For $n = 1$, we have $A = \begin{pmatrix} x_1 + y_1 \end{pmatrix}$ and thus $\det A = x_1 + y_1$.

For $n = 2$, we have $A = \begin{pmatrix} x_1 + y_1 & x_1 + y_2 \\ x_2 + y_1 & x_2 + y_2 \end{pmatrix}$ and thus $\det A = (x_1 + y_1)(x_2 + y_2) - (x_1 + y_2)(x_2 + y_1) = -(y_1 - y_2)(x_1 - x_2)$.

However, it turns out that for every $n \geq 3$, we again have $\det A = 0$. This is harder to prove than the similar claim in Example 6.6. We will eventually see how to do it easily, but as for now let me outline a direct proof. (I am being rather telegraphic here; do not worry if you do not understand the following argument, as there will be easier and more detailed proofs below.)

From (340), we obtain

$$\det A = \sum_{\sigma \in S_n} (-1)^\sigma \left(x_1 + y_{\sigma(1)}\right)\left(x_2 + y_{\sigma(2)}\right) \cdots \left(x_n + y_{\sigma(n)}\right). \tag{345}$$

If we expand the product $\left(x_1 + y_{\sigma(1)}\right)\left(x_2 + y_{\sigma(2)}\right) \cdots \left(x_n + y_{\sigma(n)}\right)$, we obtain a sum of 2^n terms:

$$\begin{aligned} &\left(x_1 + y_{\sigma(1)}\right)\left(x_2 + y_{\sigma(2)}\right) \cdots \left(x_n + y_{\sigma(n)}\right) \\ &= \sum_{I \subseteq [n]} \left(\prod_{i \in I} x_i\right)\left(\prod_{i \in [n] \setminus I} y_{\sigma(i)}\right) \end{aligned} \tag{346}$$

(where $[n]$ means the set $\{1, 2, \ldots, n\}$). (To obtain a fully rigorous proof of (346), apply Exercise 6.1 **(a)** to $a_i = x_i$ and $b_i = y_{\sigma(i)}$.) Thus, (345) becomes

$$\det A = \sum_{\sigma \in S_n} (-1)^{\sigma} \underbrace{\left(x_1 + y_{\sigma(1)}\right)\left(x_2 + y_{\sigma(2)}\right)\cdots\left(x_n + y_{\sigma(n)}\right)}_{\substack{= \sum_{I \subseteq [n]} \left(\prod_{i \in I} x_i\right)\left(\prod_{i \in [n] \setminus I} y_{\sigma(i)}\right) \\ \text{(by (346))}}}$$

$$= \sum_{\sigma \in S_n} (-1)^{\sigma} \sum_{I \subseteq [n]} \left(\prod_{i \in I} x_i\right)\left(\prod_{i \in [n] \setminus I} y_{\sigma(i)}\right)$$

$$= \sum_{I \subseteq [n]} \sum_{\sigma \in S_n} (-1)^{\sigma} \left(\prod_{i \in I} x_i\right)\left(\prod_{i \in [n] \setminus I} y_{\sigma(i)}\right).$$

We want to prove that this is 0. In order to do so, it clearly suffices to show that every $I \subseteq [n]$ satisfies

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\prod_{i \in I} x_i\right)\left(\prod_{i \in [n] \setminus I} y_{\sigma(i)}\right) = 0. \tag{347}$$

So let us fix $I \subseteq [n]$, and try to prove (347). We must be in one of the following two cases:

Case 1: The set $[n] \setminus I$ has at least two elements. In this case, let us pick two distinct elements a and b of this set, and split the set S_n into disjoint two-element subsets by pairing up every even permutation $\sigma \in S_n$ with the odd permutation $\sigma \circ t_{a,b}$ (where $t_{a,b}$ is as defined in Definition 5.29). The addends on the left hand side of (347) corresponding to two permutations paired up cancel out each other (because the products $\prod_{i \in [n] \setminus I} y_{\sigma(i)}$ and $\prod_{i \in [n] \setminus I} y_{\left(\sigma \circ t_{a,b}\right)(i)}$ differ only in the order of their factors), and thus the whole left hand side of (347) is 0. Thus, (347) is proven in Case 1.

Case 2: The set $[n] \setminus I$ has at most one element. In this case, the set I has at least two elements (it is here that we use $n \geq 3$). Pick two distinct elements c and d of I, and split the set S_n into disjoint two-element subsets by pairing up every even permutation $\sigma \in S_n$ with the odd permutation $\sigma \circ t_{c,d}$. Again, the addends on the left hand side of (347) corresponding to two permutations paired up cancel out each other (because the products $\prod_{i \in [n] \setminus I} y_{\sigma(i)}$ and $\prod_{i \in [n] \setminus I} y_{\left(\sigma \circ t_{c,d}\right)(i)}$ are identical), and thus the whole left hand side of (347) is 0. This proves (347) in Case 2.

We thus have proven (347) in both cases. So $\det A = 0$ is proven. This was a tricky argument, and shows the limits of the usefulness of (340).

We shall now discuss basic properties of the determinant.

Exercise 6.3. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n \times n$-matrix. Assume that $a_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$. Show that

$$\det A = a_{1,1}a_{2,2}\cdots a_{n,n}.$$

Definition 6.8. An $n \times n$-matrix A satisfying the assumption of Exercise 6.3 is said to be *lower-triangular* (because its entries above the diagonal are 0, and thus its nonzero entries are concentrated in the triangular region southwest of the diagonal). Exercise 6.3 thus says that the determinant of a lower-triangular matrix is the product of its diagonal entries. For instance, $\det \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix} = acf$.

Example 6.9. Let $n \in \mathbb{N}$. The $n \times n$ identity matrix I_n is lower-triangular, and its diagonal entries are $1, 1, \ldots, 1$. Hence, Exercise 6.3 shows that its determinant is $\det(I_n) = 1 \cdot 1 \cdot \cdots \cdot 1 = 1$.

Definition 6.10. The *transpose* of a matrix $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ is defined to be the matrix $\left(a_{j,i}\right)_{1\leq i\leq m,\ 1\leq j\leq n}$. It is denoted by A^T. For instance, $\begin{pmatrix} 1 & 2 & -1 \\ 4 & 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 0 \\ -1 & 1 \end{pmatrix}$.

Remark 6.11. Various other notations for the transpose of a matrix A exist in the literature. Some of them are A^t (with a lower case t) and ${}^TA$ and ${}^tA$.

Exercise 6.4. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Show that $\det\left(A^T\right) = \det A$.

The transpose of a lower-triangular $n \times n$-matrix is an upper-triangular $n \times n$-matrix (i.e., an $n \times n$-matrix whose entries below the diagonal are 0). Thus, combining Exercise 6.3 with Exercise 6.4, we see that the determinant of an upper-triangular matrix is the product of its diagonal entries.

The following exercise presents five fundamental (and simple) properties of transposes:

Exercise 6.5. Prove the following:

(a) If u, v and w are three nonnegative integers, if P is a $u \times v$-matrix, and if Q is a $v \times w$-matrix, then

$$(PQ)^T = Q^T P^T. \tag{348}$$

(b) Every $u \in \mathbb{N}$ satisfies

$$(I_u)^T = I_u. \tag{349}$$

(c) If u and v are two nonnegative integers, if P is a $u \times v$-matrix, and if $\lambda \in \mathbb{K}$, then

$$(\lambda P)^T = \lambda P^T. \tag{350}$$

(d) If u and v are two nonnegative integers, and if P and Q are two $u \times v$-matrices, then

$$(P + Q)^T = P^T + Q^T.$$

(e) If u and v are two nonnegative integers, and if P is a $u \times v$-matrix, then

$$\left(P^T\right)^T = P. \tag{351}$$

Here is yet another simple property of determinants that follows directly from their definition:

Proposition 6.12. Let $n \in \mathbb{N}$ and $\lambda \in \mathbb{K}$. Let A be an $n \times n$-matrix. Then, $\det(\lambda A) = \lambda^n \det A$.

Proof of Proposition 6.12. Write A in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus, $\lambda A = (\lambda a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of λA). Hence, (341) (applied to λA and $\lambda a_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det(\lambda A) = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^{n} \left(\lambda a_{i,\sigma(i)}\right)}_{= \lambda^n \prod_{i=1}^{n} a_{i,\sigma(i)}} = \sum_{\sigma \in S_n} (-1)^\sigma \lambda^n \prod_{i=1}^{n} a_{i,\sigma(i)}$$

$$= \lambda^n \underbrace{\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{= \det A \\ \text{(by (341))}}} = \lambda^n \det A.$$

Proposition 6.12 is thus proven. □

Exercise 6.6. Let $a, b, c, d, e, f, g, h, i, j, k, \ell, m, n, o, p$ be elements of $\mathbb{K}$.

(a) Find a simple formula for the determinant

$$\det \begin{pmatrix} a & b & c & d \\ \ell & 0 & 0 & e \\ k & 0 & 0 & f \\ j & i & h & g \end{pmatrix}.$$

(b) Find a simple formula for the determinant

$$\det \begin{pmatrix} a & b & c & d & e \\ f & 0 & 0 & 0 & g \\ h & 0 & 0 & 0 & i \\ j & 0 & 0 & 0 & k \\ \ell & m & n & o & p \end{pmatrix}.$$

(Do not mistake the "o" for a "0".)

[**Hint:** Part **(b)** is simpler than part **(a)**.]

In the next exercises, we shall talk about rows and columns; let us first make some pedantic remarks about these notions.

If $n \in \mathbb{N}$, then an $n \times 1$-matrix is said to be a *column vector* with n entries[210], whereas a $1 \times n$-matrix is said to be a *row vector* with n entries. Column vectors and row vectors store exactly the same kind of data (namely, n elements of $\mathbb{K}$), so you might wonder why I make a difference between them (and also why I distinguish them from n-tuples of elements of $\mathbb{K}$, which also contain precisely the same kind of data). The reason for this is that column vectors and row vectors behave differently under matrix multiplication: For example,

$$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix} = \begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix}$$

is not the same as

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} ac + bd \end{pmatrix}.$$

If we would identify column vectors with row vectors, then this would cause contradictions.

The reason to distinguish between row vectors and n-tuples is subtler: We have defined row vectors only for a commutative ring $\mathbb{K}$, whereas n-tuples can be made out of elements of any set. As a consequence, the sum of two row vectors is well-defined (since row vectors are matrices and thus can be added entry by entry), whereas the sum of two n-tuples is not. Similarly, we can take the product λv of an element $\lambda \in \mathbb{K}$ with a row vector v (by multiplying every entry of v by λ), but

[210] It is also called a *column vector* of size n.

such a thing does not make sense for general n-tuples. These differences between row vectors and n-tuples, however, cause no clashes of notation if we use the same notations for both types of object. Thus, we are often going to identify a row vector $\begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$ with the n-tuple $(a_1, a_2, \ldots, a_n) \in \mathbb{K}^n$. Thus, $\mathbb{K}^n$ becomes the set of all row vectors with n entries. [211]

The column vectors with n entries are in 1-to-1 correspondence with the row vectors with n entries, and this correspondence is given by taking the transpose: The column vector v corresponds to the row vector v^T, and conversely, the row vector w corresponds to the column vector w^T. In particular, every column vector $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ can be rewritten in the form $\begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}^T = (a_1, a_2, \ldots, a_n)^T$. We shall often write it in the latter form, just because it takes up less space on paper.

The rows of a matrix are row vectors; the columns of a matrix are column vectors. Thus, terms like "the sum of two rows of a matrix A" or "-3 times a column of a matrix A" make sense: Rows and columns are vectors, and thus can be added (when they have the same number of entries) and multiplied by elements of $\mathbb{K}$.

Let $n \in \mathbb{N}$ and $j \in \{1, 2, \ldots, n\}$. If v is a column vector with n entries (that is, an $n \times 1$-matrix), then the *j-th entry of v* means the $(j, 1)$-th entry of v. If v is a row vector with n entries (that is, a $1 \times n$-matrix), then the *j-th entry of v* means the $(1, j)$-th entry of v. For example, the 2-nd entry of the row vector $\begin{pmatrix} a & b & c \end{pmatrix}$ is b.

Exercise 6.7. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Prove the following:

(a) If B is an $n \times n$-matrix obtained from A by switching two rows, then $\det B = -\det A$. ("Switching two rows" means "switching two distinct rows", of course.)

(b) If B is an $n \times n$-matrix obtained from A by switching two columns, then $\det B = -\det A$.

(c) If a row of A consists of zeroes, then $\det A = 0$.

(d) If a column of A consists of zeroes, then $\det A = 0$.

(e) If A has two equal rows, then $\det A = 0$.

(f) If A has two equal columns, then $\det A = 0$.

(g) Let $\lambda \in \mathbb{K}$ and $k \in \{1, 2, \ldots, n\}$. If B is the $n \times n$-matrix obtained from A by multiplying the k-th row by λ (that is, multiplying every entry of the k-th row by λ), then $\det B = \lambda \det A$.

(h) Let $\lambda \in \mathbb{K}$ and $k \in \{1, 2, \ldots, n\}$. If B is the $n \times n$-matrix obtained from A by multiplying the k-th column by λ, then $\det B = \lambda \det A$.

(i) Let $k \in \{1, 2, \ldots, n\}$. Let A' be an $n \times n$-matrix whose rows equal the corresponding rows of A except (perhaps) the k-th row. Let B be the $n \times n$-matrix obtained from A by adding the k-th row of A' to the k-th row of A (that

[211] Some algebraists, instead, identify column vectors with n-tuples, so that $\mathbb{K}^n$ is then the set of all column vectors with n entries. This is a valid convention as well, but one must be careful not to use it simultaneously with the other convention (i.e., with the convention that row vectors are identified with n-tuples); this is why we will not use it.

is, by adding every entry of the k-th row of A' to the corresponding entry of the k-th row of A). Then, $\det B = \det A + \det A'$.

(j) Let $k \in \{1, 2, \ldots, n\}$. Let A' be an $n \times n$-matrix whose columns equal the corresponding columns of A except (perhaps) the k-th column. Let B be the $n \times n$-matrix obtained from A by adding the k-th column of A' to the k-th column of A. Then, $\det B = \det A + \det A'$.

Example 6.13. Let us show examples for several parts of Exercise 6.7 (especially, for Exercise 6.7 **(i)**, which has a somewhat daunting statement).

(a) Exercise 6.7 **(a)** yields (among other things) that

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = -\det \begin{pmatrix} g & h & i \\ d & e & f \\ a & b & c \end{pmatrix}$$

for any $a, b, c, d, e, f, g, h, i \in \mathbb{K}$.

(c) Exercise 6.7 **(c)** yields (among other things) that

$$\det \begin{pmatrix} a & b & c \\ 0 & 0 & 0 \\ d & e & f \end{pmatrix} = 0$$

for any $a, b, c, d, e, f \in \mathbb{K}$.

(e) Exercise 6.7 **(e)** yields (among other things) that

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ d & e & f \end{pmatrix} = 0$$

for any $a, b, c, d, e, f \in \mathbb{K}$.

(g) Exercise 6.7 **(g)** (applied to $n = 3$ and $k = 2$) yields that

$$\det \begin{pmatrix} a & b & c \\ \lambda d & \lambda e & \lambda f \\ g & h & i \end{pmatrix} = \lambda \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

for any $a, b, c, d, e, f \in \mathbb{K}$ and $\lambda \in \mathbb{K}$.

(i) Set $n = 3$ and $k = 2$. Set $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$. Then, a matrix A' satisfying the conditions of Exercise 6.7 **(i)** has the form $A' = \begin{pmatrix} a & b & c \\ d' & e' & f' \\ g & h & i \end{pmatrix}$. For such a matrix A', we obtain $B = \begin{pmatrix} a & b & c \\ d + d' & e + e' & f + f' \\ g & h & i \end{pmatrix}$. Exercise 6.7 **(i)** then claims that $\det B = \det A + \det A'$.

Parts **(a)**, **(c)**, **(e)**, **(g)** and **(i)** of Exercise 6.7 are often united under the slogan "the determinant of a matrix is multilinear and alternating in its rows"[212]. Similarly, parts **(b)**, **(d)**, **(f)**, **(h)** and **(j)** are combined under the slogan "the determinant of a matrix is multilinear and alternating in its columns". Many texts on linear algebra (for example, [HofKun71]) use these properties as the **definition** of the determinant[213]; this is a valid approach, but I prefer to use Definition 6.5 instead, since it is more explicit.

Exercise 6.8. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Prove the following:

(a) If we add a scalar multiple of a row of A to another row of A, then the determinant of A does not change. (A *scalar multiple* of a row vector v means a row vector of the form λv, where $\lambda \in \mathbb{K}$.)

(b) If we add a scalar multiple of a column of A to another column of A, then the determinant of A does not change. (A *scalar multiple* of a column vector v means a column vector of the form λv, where $\lambda \in \mathbb{K}$.)

Example 6.14. Let us visualize Exercise 6.8 **(a)**. Set $n = 3$ and $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$. If we add -2 times the second row of A to the first row of A, then we obtain the matrix $\begin{pmatrix} a + (-2) d & b + (-2) e & c + (-2) f \\ d & e & f \\ g & h & i \end{pmatrix}$. Exercise 6.8 **(a)** now claims that this new matrix has the same determinant as A (because -2 times the second row of A is a scalar multiple of the second row of A).

Notice the word "another" in Exercise 6.8. Adding a scalar multiple of a row of A to **the same** row of A will likely change the determinant.

[212] Specifically, parts **(c)**, **(g)** and **(i)** say that it is "multilinear", while parts **(a)** and **(e)** are responsible for the "alternating".

[213] More precisely, they define a *determinant function* to be a function $F : \mathbb{K}^{n \times n} \to \mathbb{K}$ which is multilinear and alternating in the rows of a matrix (i.e., which satisfies parts **(a)**, **(c)**, **(e)**, **(g)** and **(i)** of Exercise 6.7 if every appearance of "det" is replaced by "F" in this Exercise) and which satisfies $F(I_n) = 1$. Then, they show that there is (for each $n \in \mathbb{N}$) exactly one determinant function $F : \mathbb{K}^{n \times n} \to \mathbb{K}$. They then denote this function by det. This is a rather slick definition of a determinant, but it has the downside that it requires showing that there is exactly one determinant function (which is often not easier than our approach).

Remark 6.15. Exercise 6.8 lets us prove the claim of Example 6.7 in a much simpler way.

Namely, let n and $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ and A be as in Example 6.7. Assume that $n \geq 3$. We want to show that $\det A = 0$.

The matrix A has at least three rows (since $n \geq 3$), and looks as follows:

$$A = \begin{pmatrix} x_1 + y_1 & x_1 + y_2 & x_1 + y_3 & x_1 + y_4 & \cdots & x_1 + y_n \\ x_2 + y_1 & x_2 + y_2 & x_2 + y_3 & x_2 + y_4 & \cdots & x_2 + y_n \\ x_3 + y_1 & x_3 + y_2 & x_3 + y_3 & x_3 + y_4 & \cdots & x_3 + y_n \\ x_4 + y_1 & x_4 + y_2 & x_4 + y_3 & x_4 + y_4 & \cdots & x_4 + y_n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n + y_1 & x_n + y_2 & x_n + y_3 & x_n + y_4 & \cdots & x_n + y_n \end{pmatrix}$$

(where the presence of terms like x_4 and y_4 does not mean that the variables x_4 and y_4 exist, in the same way as one can write "$x_1, x_2, \ldots, x_k$" even if $k = 1$ or $k = 0$). Thus, if we subtract the first row of A from the second row of A, then we obtain the matrix

$$A' = \begin{pmatrix} x_1 + y_1 & x_1 + y_2 & x_1 + y_3 & x_1 + y_4 & \cdots & x_1 + y_n \\ x_2 - x_1 & x_2 - x_1 & x_2 - x_1 & x_2 - x_1 & \cdots & x_2 - x_1 \\ x_3 + y_1 & x_3 + y_2 & x_3 + y_3 & x_3 + y_4 & \cdots & x_3 + y_n \\ x_4 + y_1 & x_4 + y_2 & x_4 + y_3 & x_4 + y_4 & \cdots & x_4 + y_n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n + y_1 & x_n + y_2 & x_n + y_3 & x_n + y_4 & \cdots & x_n + y_n \end{pmatrix}$$

(because $(x_2 + y_j) - (x_1 + y_j) = x_2 - x_1$ for every j). The transformation we just did (subtracting a row from another row) does not change the determinant of the matrix (by Exercise 6.8 **(a)**, because subtracting a row from another row is tantamount to adding the (-1)-multiple of the former row to the latter), and thus we have $\det A' = \det A$.

We notice that each entry of the second row of A' equals $x_2 - x_1$.

Next, we subtract the first row of A' from the third row of A', and obtain the matrix

$$A'' = \begin{pmatrix} x_1 + y_1 & x_1 + y_2 & x_1 + y_3 & x_1 + y_4 & \cdots & x_1 + y_n \\ x_2 - x_1 & x_2 - x_1 & x_2 - x_1 & x_2 - x_1 & \cdots & x_2 - x_1 \\ x_3 - x_1 & x_3 - x_1 & x_3 - x_1 & x_3 - x_1 & \cdots & x_3 - x_1 \\ x_4 + y_1 & x_4 + y_2 & x_4 + y_3 & x_4 + y_4 & \cdots & x_4 + y_n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n + y_1 & x_n + y_2 & x_n + y_3 & x_n + y_4 & \cdots & x_n + y_n \end{pmatrix}.$$

Again, the determinant is unchanged (because of Exercise 6.8 **(a)**), so we have $\det A'' = \det A' = \det A$.

We notice that each entry of the second row of A'' equals $x_2 - x_1$ (indeed, these entries have been copied over from A'), and that each entry of the third row of A'' equals $x_3 - x_1$.

Next, we subtract the first column of A'' from each of the other columns of A''. This gives us the matrix

$$A''' = \begin{pmatrix} x_1 + y_1 & y_2 - y_1 & y_3 - y_1 & y_4 - y_1 & \cdots & y_n - y_1 \\ x_2 - x_1 & 0 & 0 & 0 & \cdots & 0 \\ x_3 - x_1 & 0 & 0 & 0 & \cdots & 0 \\ x_4 + y_1 & y_2 - y_1 & y_3 - y_1 & y_4 - y_1 & \cdots & y_n - y_1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n + y_1 & y_2 - y_1 & y_3 - y_1 & y_4 - y_1 & \cdots & y_n - y_1 \end{pmatrix}. \tag{352}$$

This step, again, has not changed the determinant (because Exercise 6.8 **(b)** shows that subtracting a column from another column does not change the determinant, and what we did was doing $n-1$ such transformations). Thus, $\det A''' = \det A'' = \det A$.

Now, let us write the matrix A''' in the form $A''' = \left(a'''_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. (Thus, $a'''_{i,j}$ is the (i,j)-th entry of A''' for every (i,j).) Then, (340) (applied to A''' instead of A) yields

$$\det A''' = \sum_{\sigma\in S_n} (-1)^\sigma a'''_{1,\sigma(1)} a'''_{2,\sigma(2)} \cdots a'''_{n,\sigma(n)}. \tag{353}$$

I claim that

$$a'''_{1,\sigma(1)} a'''_{2,\sigma(2)} \cdots a'''_{n,\sigma(n)} = 0 \qquad \text{for every } \sigma \in S_n. \tag{354}$$

[*Proof of (354):* Let $\sigma \in S_n$. Then, σ is injective, and thus $\sigma(2) \neq \sigma(3)$. Therefore, at least one of the integers $\sigma(2)$ and $\sigma(3)$ must be $\neq 1$ (because otherwise, we would have $\sigma(2) = 1 = \sigma(3)$, contradicting $\sigma(2) \neq \sigma(3)$). We WLOG assume that $\sigma(2) \neq 1$. But a look at (352) reveals that all entries of the second row of A''' are zero except for the first entry. Thus, $a'''_{2,j} = 0$ for every $j \neq 1$. Applied to $j = \sigma(2)$, this yields $a'''_{2,\sigma(2)} = 0$ (since $\sigma(2) \neq 1$). Hence, $a'''_{1,\sigma(1)} a'''_{2,\sigma(2)} \cdots a'''_{n,\sigma(n)} = 0$ (because if 0 appears as a factor in a product, then the whole product must be 0). This proves (354).]

Now, (353) becomes

$$\det A''' = \sum_{\sigma\in S_n} (-1)^\sigma \underbrace{a'''_{1,\sigma(1)} a'''_{2,\sigma(2)} \cdots a'''_{n,\sigma(n)}}_{\substack{=0 \\ \text{(by (354))}}} = \sum_{\sigma\in S_n} (-1)^\sigma 0 = 0.$$

Compared with $\det A''' = \det A$, this yields $\det A = 0$. Thus, $\det A = 0$ is proven again.

Remark 6.16. Here is another example for the use of Exercise 6.8.

Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let A be the matrix $\left(x_{\max\{i,j\}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. (Recall that $\max S$ denotes the greatest element of a nonempty set S.)

For example, if $n = 4$, then

$$A = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_2 & x_3 & x_4 \\ x_3 & x_3 & x_3 & x_4 \\ x_4 & x_4 & x_4 & x_4 \end{pmatrix}.$$

We want to find $\det A$. First, let us subtract the first row of A from each of the other rows of A. Thus we obtain a new matrix A'. The determinant has not changed (according to Exercise 6.8 **(a)**); i.e., we have $\det A' = \det A$. Here is how A' looks like in the case when $n = 4$:

$$A' = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 - x_1 & 0 & 0 & 0 \\ x_3 - x_1 & x_3 - x_2 & 0 & 0 \\ x_4 - x_1 & x_4 - x_2 & x_4 - x_3 & 0 \end{pmatrix}. \tag{355}$$

Notice the many zeroes; zeroes are useful when computing determinants. To generalize the pattern we see on (355), we write the matrix A' in the form $A' = \left(a'_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (so that $a'_{i,j}$ is the (i,j)-th entry of A' for every (i,j)). Then, for every $(i,j) \in \{1, 2, \ldots, n\}^2$, we have

$$a'_{i,j} = \begin{cases} x_{\max\{i,j\}}, & \text{if } i = 1; \\ x_{\max\{i,j\}} - x_{\max\{1,j\}}, & \text{if } i > 1 \end{cases} \tag{356}$$

(since we obtained the matrix A' by subtracting the first row of A from each of the other rows of A). Hence, for every $(i,j) \in \{1, 2, \ldots, n\}^2$ satisfying $1 < i \leq j$, we have

$$a'_{i,j} = x_{\max\{i,j\}} - x_{\max\{1,j\}} = x_j - x_j \qquad \left(\begin{array}{c} \text{since } \max\{i,j\} = j \text{ (because } i \leq j) \\ \text{and } \max\{1,j\} = j \text{ (because } 1 < j) \end{array}\right)$$
$$= 0. \tag{357}$$

This is the general explanation for the six 0's in (355). We notice also that the first row of the matrix A' is $(x_1, x_2, \ldots, x_n)$.

Now, we want to transform A' further. Namely, we first switch the first row with the second row; then we switch the second row (which used to be the first row) with the third row; then, the third row with the fourth row, and so on, until we finally switch the $(n-1)$-th row with the n-th row. As a result of these $n-1$

switches, the first row has moved all the way down to the bottom, past all the other rows. We denote the resulting matrix by A''. For instance, if $n = 4$, then

$$A'' = \begin{pmatrix} x_2 - x_1 & 0 & 0 & 0 \\ x_3 - x_1 & x_3 - x_2 & 0 & 0 \\ x_4 - x_1 & x_4 - x_2 & x_4 - x_3 & 0 \\ x_1 & x_2 & x_3 & x_4 \end{pmatrix}. \tag{358}$$

This is a lower-triangular matrix. To see that this holds in the general case, we write the matrix A'' in the form $A'' = \left(a''_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (so that $a''_{i,j}$ is the (i,j)-th entry of A'' for every (i,j)). Then, for every $(i,j) \in \{1,2,\ldots,n\}^2$, we have

$$a''_{i,j} = \begin{cases} a'_{i+1,j}, & \text{if } i < n; \\ a'_{1,j}, & \text{if } i = n \end{cases} \tag{359}$$

(because the first row of A' has become the n-th row of A'', whereas every other row has moved up one step). In particular, for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $1 \leq i < j \leq n$, we have

$$\begin{aligned} a''_{i,j} &= \begin{cases} a'_{i+1,j}, & \text{if } i < n; \\ a'_{1,j}, & \text{if } i = n \end{cases} = a'_{i+1,j} \qquad (\text{since } i < j \leq n) \\ &= 0 \qquad \left(\begin{array}{c} \text{by (357), applied to } i+1 \text{ instead of } i \\ (\text{because } i < j \text{ yields } i+1 \leq j) \end{array} \right). \end{aligned}$$

This shows that A'' is indeed lower-triangular. Hence, Exercise 6.3 (applied to A'' and $a''_{i,j}$ instead of A and $a_{i,j}$) shows that $\det A'' = a''_{1,1} a''_{2,2} \cdots a''_{n,n}$.

Using (359) and (356), it is easy to see that every $i \in \{1,2,\ldots,n\}$ satisfies

$$a''_{i,i} = \begin{cases} x_{i+1} - x_i, & \text{if } i < n; \\ x_n, & \text{if } i = n \end{cases}. \tag{360}$$

(This is precisely the pattern you would guess from the diagonal entries in (358).) Now, multiplying the equalities (360) for all $i \in \{1,2,\ldots,n\}$, we obtain $a''_{1,1} a''_{2,2} \cdots a''_{n,n} = (x_2 - x_1)(x_3 - x_2) \cdots (x_n - x_{n-1}) x_n$. Thus,

$$\det A'' = a''_{1,1} a''_{2,2} \cdots a''_{n,n} = (x_2 - x_1)(x_3 - x_2) \cdots (x_n - x_{n-1}) x_n. \tag{361}$$

But we want $\det A$, not $\det A''$. First, let us find $\det A'$. Recall that A'' was obtained from A' by switching rows, repeatedly – namely, $n-1$ times. Every time we switch two rows in a matrix, its determinant gets multiplied by -1 (because of Exercise 6.7 **(a)**). Hence, $n-1$ such switches cause the determinant to be multiplied by $(-1)^{n-1}$. Since A'' was obtained from A' by $n-1$ such switches, we thus conclude that $\det A'' = (-1)^{n-1} \det A'$, so that

$$\det A' = \underbrace{\frac{1}{(-1)^{n-1}}}_{=(-1)^{n-1}} \underbrace{\det A''}_{=(x_2-x_1)(x_3-x_2)\cdots(x_n-x_{n-1})x_n} = (-1)^{n-1} (x_2 - x_1)(x_3 - x_2) \cdots (x_n - x_{n-1}) x_n.$$

Finally, recall that $\det A' = \det A$, so that

$$\det A = \det A' = (-1)^{n-1} (x_2 - x_1)(x_3 - x_2) \cdots (x_n - x_{n-1}) x_n.$$

6.4. $\det(AB)$

Next, a lemma that will come handy in a more important proof:

Lemma 6.17. Let $n \in \mathbb{N}$. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Let $\kappa : [n] \to [n]$ be a map. Let $B = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. Let B_κ be the $n \times n$-matrix $\left(b_{\kappa(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

(a) If $\kappa \in S_n$, then $\det(B_\kappa) = (-1)^\kappa \cdot \det B$.

(b) If $\kappa \notin S_n$, then $\det(B_\kappa) = 0$.

Remark 6.18. Lemma 6.17 **(a)** simply says that if we permute the rows of a square matrix, then its determinant gets multiplied by the sign of the permutation used. For instance, let $n = 3$ and $B = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$. If κ is the permutation $(2, 3, 1)$ (in one-line notation), then $B_\kappa = \begin{pmatrix} d & e & f \\ g & h & i \\ a & b & c \end{pmatrix}$, and Lemma 6.17 **(a)** says that $\det(B_\kappa) = \underbrace{(-1)^\kappa}_{=1} \cdot \det B = \det B$.

Of course, a similar result holds for permutations of columns.

Remark 6.19. Exercise 6.7 **(a)** is a particular case of Lemma 6.17 **(a)**. Indeed, if B is an $n \times n$-matrix obtained from A by switching the u-th and the v-th row (where u and v are two distinct elements of $\{1, 2, \ldots, n\}$), then $B = \left(a_{t_{u,v}(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (where A is written in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$).

Proof of Lemma 6.17. Recall that S_n is the set of all permutations of $\{1, 2, \ldots, n\}$. In other words, S_n is the set of all permutations of $[n]$ (since $[n] = \{1, 2, \ldots, n\}$). In other words, S_n is the set of all bijective maps $[n] \to [n]$.

(a) Assume that $\kappa \in S_n$. We define a map $\Phi : S_n \to S_n$ by

$$\Phi(\sigma) = \sigma \circ \kappa \qquad \text{for every } \sigma \in S_n.$$

We also define a map $\Psi : S_n \to S_n$ by

$$\Psi(\sigma) = \sigma \circ \kappa^{-1} \qquad \text{for every } \sigma \in S_n.$$

The maps Φ and Ψ are mutually inverse[214]. Hence, the map Φ is a bijection.

We have $B = (b_{i,j})_{1 \le i \le n,\ 1 \le j \le n}$. Hence, (341) (applied to B and $b_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det B = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^{n}}_{= \prod_{i \in [n]}} b_{i,\sigma(i)} = \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i \in [n]} b_{i,\sigma(i)}. \tag{362}$$

Now, $B_\kappa = \left(b_{\kappa(i),j}\right)_{1 \le i \le n,\ 1 \le j \le n}$. Hence, (341) (applied to B_κ and $b_{\kappa(i),j}$ instead of A and $a_{i,j}$) yields

$$\begin{aligned} \det(B_\kappa) &= \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^{n}}_{= \prod_{i \in [n]}} b_{\kappa(i),\sigma(i)} = \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i \in [n]} b_{\kappa(i),\sigma(i)} \\ &= \sum_{\sigma \in S_n} (-1)^{\Phi(\sigma)} \prod_{i \in [n]} b_{\kappa(i),(\Phi(\sigma))(i)} \end{aligned} \tag{363}$$

(here, we have substituted $\Phi(\sigma)$ for σ in the sum, since Φ is a bijection).

But every $\sigma \in S_n$ satisfies $(-1)^{\Phi(\sigma)} = (-1)^\kappa \cdot (-1)^\sigma$ [215] and $\prod_{i \in [n]} b_{\kappa(i),(\Phi(\sigma))(i)} =$

---

[214] *Proof.* Every $\sigma \in S_n$ satisfies

$$\begin{aligned} (\Psi \circ \Phi)(\sigma) &= \Psi\left(\underbrace{\Phi(\sigma)}_{=\sigma \circ \kappa}\right) = \Psi(\sigma \circ \kappa) = \sigma \circ \underbrace{\kappa \circ \kappa^{-1}}_{=\mathrm{id}} \qquad (\text{by the definition of } \Psi) \\ &= \sigma = \mathrm{id}(\sigma). \end{aligned}$$

Thus, $\Psi \circ \Phi = \mathrm{id}$. Similarly, $\Phi \circ \Psi = \mathrm{id}$. Combined with $\Psi \circ \Phi = \mathrm{id}$, this yields that the maps Φ and Ψ are mutually inverse, qed.

[215] *Proof.* Let $\sigma \in S_n$. Then, $\Phi(\sigma) = \sigma \circ \kappa$, so that

$$\begin{aligned} (-1)^{\Phi(\sigma)} &= (-1)^{\sigma \circ \kappa} = (-1)^\sigma \cdot (-1)^\kappa \qquad (\text{by (315), applied to } \tau = \kappa) \\ &= (-1)^\kappa \cdot (-1)^\sigma, \end{aligned}$$

qed.

$\prod_{i\in[n]} b_{i,\sigma(i)}$ [216]. Thus, (363) becomes

$$\det(B_\kappa) = \sum_{\sigma\in S_n} \underbrace{(-1)^{\Phi(\sigma)}}_{=(-1)^\kappa\cdot(-1)^\sigma} \underbrace{\prod_{i\in[n]} b_{\kappa(i),(\Phi(\sigma))(i)}}_{=\prod_{i\in[n]} b_{i,\sigma(i)}} = \sum_{\sigma\in S_n} (-1)^\kappa \cdot (-1)^\sigma \prod_{i\in[n]} b_{i,\sigma(i)}$$
$$= (-1)^\kappa \cdot \underbrace{\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i\in[n]} b_{i,\sigma(i)}}_{\substack{=\det B \\ \text{(by (362))}}} = (-1)^\kappa \cdot \det B.$$

This proves Lemma 6.17 **(a)**.

(b) Assume that $\kappa \notin S_n$.

The following fact is well-known: If U is a finite set, then every injective map $U \to U$ is bijective[217]. We can apply this to $U = [n]$, and thus conclude that every injective map $[n] \to [n]$ is bijective. Therefore, if the map $\kappa : [n] \to [n]$ were injective, then κ would be bijective and therefore would be an element of S_n (since S_n is the set of all bijective maps $[n] \to [n]$); but this would contradict the fact that $\kappa \notin S_n$. Hence, the map $\kappa : [n] \to [n]$ cannot be injective. Therefore, there exist two distinct elements a and b of $[n]$ such that $\kappa(a) = \kappa(b)$. Consider these a and b.

Thus, a and b are two distinct elements of $[n] = \{1, 2, \ldots, n\}$. Hence, a transposition $t_{a,b} \in S_n$ is defined (see Definition 5.29 for the definition). This transposition satisfies $\kappa \circ t_{a,b} = \kappa$ [218]. Exercise 5.10 **(b)** (applied to $i = a$ and $j = b$) yields $(-1)^{t_{a,b}} = -1$.

---

[216] *Proof.* Let $\sigma \in S_n$. We have $\Phi(\sigma) = \sigma \circ \kappa$. Thus, for every $i \in [n]$, we have $(\Phi(\sigma))(i) = (\sigma \circ \kappa)(i) = \sigma(\kappa(i))$. Hence, $\prod_{i\in[n]} b_{\kappa(i),(\Phi(\sigma))(i)} = \prod_{i\in[n]} b_{\kappa(i),\sigma(\kappa(i))}$.

But $\kappa \in S_n$. In other words, κ is a permutation of the set $\{1, 2, \ldots, n\} = [n]$, hence a bijection from $[n]$ to $[n]$. Therefore, we can substitute $\kappa(i)$ for i in the product $\prod_{i\in[n]} b_{i,\sigma(i)}$. We thus obtain $\prod_{i\in[n]} b_{i,\sigma(i)} = \prod_{i\in[n]} b_{\kappa(i),\sigma(\kappa(i))}$. Comparing this with $\prod_{i\in[n]} b_{\kappa(i),(\Phi(\sigma))(i)} = \prod_{i\in[n]} b_{\kappa(i),\sigma(\kappa(i))}$, we obtain $\prod_{i\in[n]} b_{\kappa(i),(\Phi(\sigma))(i)} = \prod_{i\in[n]} b_{i,\sigma(i)}$, qed.

[217] *Proof.* Let U be a finite set, and let f be an injective map $U \to U$. We must show that f is bijective.

Since f is injective, we have $|f(U)| = |U|$. Thus, $f(U)$ is a subset of U which has size $|U|$. But the only such subset is U itself (since U is a finite set). Therefore, $f(U)$ must be U itself. In other words, the map f is surjective. Hence, f is bijective (since f is injective and surjective), qed.

[218] *Proof.* We are going to show that every $i \in [n]$ satisfies $(\kappa \circ t_{a,b})(i) = \kappa(i)$.

So let $i \in [n]$. We shall show that $(\kappa \circ t_{a,b})(i) = \kappa(i)$.

The definition of $t_{a,b}$ shows that $t_{a,b}$ is the permutation in S_n which switches a with b while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged. In other words, we have $t_{a,b}(a) = b$, and $t_{a,b}(b) = a$, and $t_{a,b}(j) = j$ for every $j \in [n] \setminus \{a, b\}$.

Now, we have $i \in [n]$. Thus, we are in one of the following three cases:

Case 1: We have $i = a$.

Case 2: We have $i = b$.

Case 3: We have $i \in [n] \setminus \{a, b\}$.

Let A_n be the set of all even permutations in S_n. Let C_n be the set of all odd permutations in S_n.

We have $\sigma \circ t_{a,b} \in C_n$ for every $\sigma \in A_n$ [219]. Hence, we can define a map $\Phi : A_n \to C_n$ by

$$\Phi(\sigma) = \sigma \circ t_{a,b} \qquad \text{for every } \sigma \in A_n.$$

Consider this map Φ. Furthermore, we have $\sigma \circ (t_{a,b})^{-1} \in A_n$ for every $\sigma \in C_n$ [220]. Thus, we can define a map $\Psi : C_n \to A_n$ by

$$\Psi(\sigma) = \sigma \circ (t_{a,b})^{-1} \qquad \text{for every } \sigma \in C_n.$$

Consider this map Ψ.

Let us first consider Case 1. In this case, we have $i = a$, so that $(\kappa \circ t_{a,b})\left(\underbrace{i}_{=a}\right) = (\kappa \circ t_{a,b})(a) = \kappa\left(\underbrace{t_{a,b}(a)}_{=b}\right) = \kappa(b)$. Compared with $\kappa\left(\underbrace{i}_{=a}\right) = \kappa(a) = \kappa(b)$, this yields $(\kappa \circ t_{a,b})(i) = \kappa(i)$. Thus, $(\kappa \circ t_{a,b})(i) = \kappa(i)$ is proven in Case 1.

Let us next consider Case 2. In this case, we have $i = b$, so that $(\kappa \circ t_{a,b})\left(\underbrace{i}_{=b}\right) = (\kappa \circ t_{a,b})(b) = \kappa\left(\underbrace{t_{a,b}(b)}_{=a}\right) = \kappa(a) = \kappa(b)$. Compared with $\kappa\left(\underbrace{i}_{=b}\right) = \kappa(b)$, this yields $(\kappa \circ t_{a,b})(i) = \kappa(i)$. Thus, $(\kappa \circ t_{a,b})(i) = \kappa(i)$ is proven in Case 2.

Let us finally consider Case 3. In this case, we have $i \in [n] \setminus \{a,b\}$. Hence, $t_{a,b}(i) = i$ (since $t_{a,b}(j) = j$ for every $j \in [n] \setminus \{a,b\}$). Therefore, $(\kappa \circ t_{a,b})(i) = \kappa\left(\underbrace{t_{a,b}(i)}_{=i}\right) = \kappa(i)$. Thus, $(\kappa \circ t_{a,b})(i) = \kappa(i)$ is proven in Case 3.

We now have shown $(\kappa \circ t_{a,b})(i) = \kappa(i)$ in each of the three Cases 1, 2 and 3. Hence, $(\kappa \circ t_{a,b})(i) = \kappa(i)$ always holds.

Now, let us forget that we fixed i. We thus have shown that $(\kappa \circ t_{a,b})(i) = \kappa(i)$ for every $i \in [n]$. In other words, $\kappa \circ t_{a,b} = \kappa$, qed.

[219] *Proof.* Let $\sigma \in A_n$. Then, σ is an even permutation in S_n (since A_n is the set of all even permutations in S_n). Hence, $(-1)^\sigma = 1$. Now, (315) (applied to $\tau = t_{a,b}$) yields $(-1)^{\sigma \circ t_{a,b}} = \underbrace{(-1)^\sigma}_{=1} \cdot \underbrace{(-1)^{t_{a,b}}}_{=-1} = -1$. Thus, the permutation $\sigma \circ t_{a,b}$ is odd. Hence, $\sigma \circ t_{a,b}$ is an odd permutation in S_n. In other words, $\sigma \circ t_{a,b} \in C_n$ (since C_n is the set of all odd permutations in S_n), qed.

[220] *Proof.* Let $\sigma \in C_n$. Then, σ is an odd permutation in S_n (since C_n is the set of all odd permutations in S_n). Hence, $(-1)^\sigma = -1$.

Applying (316) to $t_{a,b}$ instead of σ, we obtain $(-1)^{(t_{a,b})^{-1}} = (-1)^{t_{a,b}} = -1$. Now, (315) (applied to $\tau = (t_{a,b})^{-1}$) yields $(-1)^{\sigma \circ (t_{a,b})^{-1}} = \underbrace{(-1)^\sigma}_{=-1} \cdot \underbrace{(-1)^{(t_{a,b})^{-1}}}_{=-1} = (-1) \cdot (-1) = 1$. Thus, the permutation $\sigma \circ (t_{a,b})^{-1}$ is even. Hence, $\sigma \circ (t_{a,b})^{-1}$ is an even permutation in S_n. In other words, $\sigma \circ (t_{a,b})^{-1} \in A_n$ (since A_n is the set of all even permutations in S_n), qed.

(We could have simplified our life a bit by noticing that $\left(t_{a,b}\right)^{-1} = t_{a,b}$, so that the maps Φ and Ψ are given by the same formula, albeit defined on different domains. But I wanted to demonstrate a use of (316).)

The maps Φ and Ψ are mutually inverse[221]. Hence, the map Ψ is a bijection. Moreover, every $\sigma \in C_n$ satisfies

$$\prod_{i\in[n]} b_{\kappa(i),(\Psi(\sigma))(i)} = \prod_{i\in[n]} b_{\kappa(i),\sigma(i)}. \tag{364}$$

[222]

We have $B_\kappa = \left(b_{\kappa(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence, (341) (applied to B_κ and $b_{\kappa(i),j}$ instead

[221] *Proof.* Every $\sigma \in A_n$ satisfies

$$\begin{aligned}(\Psi\circ\Phi)(\sigma) &= \Psi\left(\underbrace{\Phi(\sigma)}_{=\sigma\circ\tau_{a,b}}\right) = \Psi\left(\sigma\circ\tau_{a,b}\right) = \sigma\circ\underbrace{\tau_{a,b}\circ\left(\tau_{a,b}\right)^{-1}}_{=\operatorname{id}} \qquad (\text{by the definition of } \Psi)\\ &= \sigma = \operatorname{id}(\sigma).\end{aligned}$$

Thus, $\Psi\circ\Phi = \operatorname{id}$. Similarly, $\Phi\circ\Psi = \operatorname{id}$. Combined with $\Psi\circ\Phi = \operatorname{id}$, this yields that the maps Φ and Ψ are mutually inverse, qed.

[222] *Proof of (364):* Let $\sigma \in C_n$. The map $t_{a,b}$ is a permutation of $[n]$, thus a bijection $[n] \to [n]$. Hence, we can substitute $t_{a,b}(i)$ for i in the product $\prod_{i\in[n]} b_{\kappa(i),(\Psi(\sigma))(i)}$. Thus we obtain

$$\prod_{i\in[n]} b_{\kappa(i),(\Psi(\sigma))(i)} = \prod_{i\in[n]} b_{\kappa\left(t_{a,b}(i)\right),(\Psi(\sigma))\left(t_{a,b}(i)\right)} = \prod_{i\in[n]} b_{\kappa(i),\sigma(i)}$$

(since every $i \in [n]$ satisfies $\kappa\left(t_{a,b}(i)\right) = \underbrace{\left(\kappa\circ t_{a,b}\right)}_{=\kappa}(i) = \kappa(i)$ and

$$\underbrace{(\Psi(\sigma))}_{=\sigma\circ\left(t_{a,b}\right)^{-1}}\left(t_{a,b}(i)\right) = \left(\sigma\circ\left(t_{a,b}\right)^{-1}\right)\left(t_{a,b}(i)\right) = \sigma\left(\underbrace{\left(t_{a,b}\right)^{-1}\left(t_{a,b}(i)\right)}_{=i}\right) = \sigma(i)$$

). This proves (364).

of A and $a_{i,j}$) yields

$$\begin{aligned}
\det(B_\kappa) &= \sum_{\sigma\in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^n}_{=\prod_{i\in[n]}} b_{\kappa(i),\sigma(i)} = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} \\
&= \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma \text{ is even}}}}_{\substack{=\sum_{\sigma\in A_n}\\ \text{(since } A_n \text{ is the}\\ \text{set of all even}\\ \text{permutations}\\ \text{in } S_n)}} \underbrace{(-1)^\sigma}_{\substack{=1\\ \text{(since } \sigma \text{ is even)}}} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} + \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma \text{ is odd}}}}_{\substack{=\sum_{\sigma\in C_n}\\ \text{(since } C_n \text{ is the}\\ \text{set of all odd}\\ \text{permutations}\\ \text{in } S_n)}} \underbrace{(-1)^\sigma}_{\substack{=-1\\ \text{(since } \sigma \text{ is odd)}}} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} \\
&\qquad (\text{since every permutation } \sigma\in S_n \text{ is either even or odd, but not both}) \\
&= \sum_{\sigma\in A_n} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} + \sum_{\sigma\in C_n} (-1) \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} \\
&= \sum_{\sigma\in A_n} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} - \sum_{\sigma\in C_n} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} = 0,
\end{aligned}$$

since

$$\begin{aligned}
&\sum_{\sigma\in A_n} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)} \\
&= \sum_{\sigma\in C_n} \underbrace{\prod_{i\in[n]} b_{\kappa(i),(\Psi(\sigma))(i)}}_{\substack{=\prod_{i\in[n]} b_{\kappa(i),\sigma(i)}\\ \text{(by (364))}}} \\
&\qquad (\text{here, we have substituted } \Psi(\sigma) \text{ for } \sigma, \text{ since the map } \Psi \text{ is a bijection}) \\
&= \sum_{\sigma\in C_n} \prod_{i\in[n]} b_{\kappa(i),\sigma(i)}.
\end{aligned}$$

This proves Lemma 6.17 **(b)**. □

Now let us state a basic formula for products of sums in a commutative ring:

Lemma 6.20. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $n \in \mathbb{N}$. For every $i \in [n]$, let $p_{i,1}, p_{i,2}, \ldots, p_{i,m_i}$ be finitely many elements of $\mathbb{K}$. Then,

$$\prod_{i=1}^n \sum_{k=1}^{m_i} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_n)\in[m_1]\times[m_2]\times\cdots\times[m_n]} \prod_{i=1}^n p_{i,k_i}.$$

(**Pedantic remark:** If $n = 0$, then the Cartesian product $[m_1] \times [m_2] \times \cdots \times [m_n]$ has no factors; it is what is called an *empty Cartesian product*. It is understood to be a 1-element set, and its single element is the 0-tuple $()$ (also known as the empty list).)

I tend to refer to Lemma 6.20 as the *product rule* (since it is related to the product rule for joint probabilities); I think it has no really widespread name. However, it is a fundamental algebraic fact that is used very often and tacitly (I suspect that most mathematicians have never thought of it as being a theorem that needs to be proven). The idea behind Lemma 6.20 is that if you expand the product

$$\begin{aligned}
&\prod_{i=1}^{n} \sum_{k=1}^{m_i} p_{i,k} \\
&= \prod_{i=1}^{n} \left(p_{i,1} + p_{i,2} + \cdots + p_{i,m_i}\right) \\
&= \left(p_{1,1} + p_{1,2} + \cdots + p_{1,m_1}\right) \left(p_{2,1} + p_{2,2} + \cdots + p_{2,m_2}\right) \cdots \left(p_{n,1} + p_{n,2} + \cdots + p_{n,m_n}\right),
\end{aligned}$$

then you get a sum of $m_1 m_2 \cdots m_n$ terms, each of which has the form

$$p_{1,k_1} p_{2,k_2} \cdots p_{n,k_n} = \prod_{i=1}^{n} p_{i,k_i}$$

for some $(k_1, k_2, \ldots, k_n) \in [m_1] \times [m_2] \times \cdots \times [m_n]$. (More precisely, it is the sum of all such terms.) A formal proof of Lemma 6.20 could be obtained by induction over n using the distributivity axiom[223]. For the details (if you care about them), see the solution to the following exercise:

Exercise 6.9. Prove Lemma 6.20.

Remark 6.21. Lemma 6.20 can be regarded as a generalization of Exercise 6.1 **(a)**. Indeed, let me sketch how Exercise 6.1 **(a)** can be derived from Lemma 6.20:

Let n, $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ be as in Exercise 6.1 **(a)**. For every $i \in [n]$, set $m_i = 2$, $p_{i,1} = a_i$ and $p_{i,2} = b_i$. Then, Lemma 6.20 yields

$$\begin{aligned}
\prod_{i=1}^{n} (a_i + b_i) &= \underbrace{\sum_{(k_1, k_2, \ldots, k_n) \in \underbrace{[2] \times [2] \times \cdots \times [2]}_{n \text{ factors}}}}_{= \sum_{(k_1, k_2, \ldots, k_n) \in [2]^n}} \underbrace{\prod_{i=1}^{n} p_{i,k_i}}_{= \left(\prod_{i \in [n];\ k_i = 1} a_i\right) \left(\prod_{i \in [n];\ k_i = 2} b_i\right)} \\
&= \sum_{(k_1, k_2, \ldots, k_n) \in [2]^n} \left(\prod_{\substack{i \in [n]; \\ k_i = 1}} a_i\right) \left(\prod_{\substack{i \in [n]; \\ k_i = 2}} b_i\right). \qquad (365)
\end{aligned}$$

[223] and the observation that the n-tuples $(k_1, k_2, \ldots, k_n) \in [m_1] \times [m_2] \times \cdots \times [m_n]$ are in bijection with the pairs $((k_1, k_2, \ldots, k_{n-1}), k_n)$ of an $(n-1)$-tuple $(k_1, k_2, \ldots, k_{n-1}) \in [m_1] \times [m_2] \times \cdots \times [m_{n-1}]$ and an element $k_n \in [m_n]$

But there is a bijection between the set $[2]^n$ and the powerset $\mathcal{P}([n])$ of $[n]$: Namely, to every n-tuple $(k_1, k_2, \ldots, k_n) \in [2]^n$, we can assign the set $\{i \in [n] \mid k_i = 1\} \in \mathcal{P}([n])$. It is easy to see that this assignment really is a bijection $[2]^n \to \mathcal{P}([n])$, and that it furthermore has the property that every n-tuple $(k_1, k_2, \ldots, k_n) \in [2]^n$ satisfies

$$\left(\prod_{\substack{i \in [n]; \\ k_i = 1}} a_i\right)\left(\prod_{\substack{i \in [n]; \\ k_i = 2}} b_i\right) = \left(\prod_{i \in I} a_i\right)\left(\prod_{i \in [n] \setminus I} b_i\right),$$

where I is the image of $(k_1, k_2, \ldots, k_n)$ under this bijection. Hence,

$$\begin{aligned}
&\sum_{(k_1, k_2, \ldots, k_n) \in [2]^n} \left(\prod_{\substack{i \in [n]; \\ k_i = 1}} a_i\right)\left(\prod_{\substack{i \in [n]; \\ k_i = 2}} b_i\right) \\
&= \sum_{I \subseteq [n]} \left(\prod_{i \in I} a_i\right)\left(\prod_{i \in [n] \setminus I} b_i\right).
\end{aligned}$$

Hence, (365) rewrites as

$$\prod_{i=1}^{n} (a_i + b_i) = \sum_{I \subseteq [n]} \left(\prod_{i \in I} a_i\right)\left(\prod_{i \in [n] \setminus I} b_i\right).$$

But this is precisely the claim of Exercise 6.1 **(a)**.

We shall use a corollary of Lemma 6.20:

Lemma 6.22. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. For every $i \in [n]$, let $p_{i,1}, p_{i,2}, \ldots, p_{i,m}$ be m elements of $\mathbb{K}$. Then,

$$\prod_{i=1}^{n} \sum_{k=1}^{m} p_{i,k} = \sum_{\kappa : [n] \to [m]} \prod_{i=1}^{n} p_{i, \kappa(i)}.$$

Proof of Lemma 6.22. For the sake of completeness, let us give this proof.

Lemma 6.20 (applied to $m_i = m$ for every $i \in [n]$) yields

$$\prod_{i=1}^{n} \sum_{k=1}^{m} p_{i,k} = \sum_{(k_1, k_2, \ldots, k_n) \in \underbrace{[m] \times [m] \times \cdots \times [m]}_{n \text{ factors}}} \prod_{i=1}^{n} p_{i, k_i}. \qquad (366)$$

Let $\operatorname{Map}([n],[m])$ denote the set of all functions from $[n]$ to $[m]$. Now, let Φ be the map from $\operatorname{Map}([n],[m])$ to $\underbrace{[m] \times [m] \times \cdots \times [m]}_{n \text{ factors}}$ given by

$$\Phi(\kappa) = (\kappa(1), \kappa(2), \ldots, \kappa(n)) \qquad \text{for every } \kappa \in \operatorname{Map}([n],[m]).$$

So the map Φ takes a function κ from $[n]$ to $[m]$, and outputs the list $(\kappa(1), \kappa(2), \ldots, \kappa(n))$ of all its values. Clearly, the map Φ is injective (since a function $\kappa \in \operatorname{Map}([n],[m])$ can be reconstructed from the list $(\kappa(1), \kappa(2), \ldots, \kappa(n)) = \Phi(\kappa)$) and surjective (since every list of n elements of $[m]$ is the list of values of some function $\kappa \in \operatorname{Map}([n],[m])$). Thus, Φ is bijective. Therefore, we can substitute $\Phi(\kappa)$ for $(k_1, k_2, \ldots, k_n)$ in the sum $\sum\limits_{(k_1,k_2,\ldots,k_n) \in \underbrace{[m] \times [m] \times \cdots \times [m]}_{n \text{ factors}}} \prod\limits_{i=1}^{n} p_{i,k_i}$.
In other words, we can substitute $(\kappa(1), \kappa(2), \ldots, \kappa(n))$ for $(k_1, k_2, \ldots, k_n)$ in this sum (since $\Phi(\kappa) = (\kappa(1), \kappa(2), \ldots, \kappa(n))$ for each $\kappa \in \operatorname{Map}([n],[m])$). We thus obtain

$$\sum_{(k_1,k_2,\ldots,k_n) \in \underbrace{[m] \times [m] \times \cdots \times [m]}_{n \text{ factors}}} \prod_{i=1}^{n} p_{i,k_i} = \underbrace{\sum_{\kappa \in \operatorname{Map}([n],[m])}}_{= \sum_{\kappa:[n]\to[m]}} \prod_{i=1}^{n} p_{i,\kappa(i)} = \sum_{\kappa:[n]\to[m]} \prod_{i=1}^{n} p_{i,\kappa(i)}.$$

Thus, (366) becomes

$$\prod_{i=1}^{n} \sum_{k=1}^{m} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_n) \in \underbrace{[m] \times [m] \times \cdots \times [m]}_{n \text{ factors}}} \prod_{i=1}^{n} p_{i,k_i} = \sum_{\kappa:[n]\to[m]} \prod_{i=1}^{n} p_{i,\kappa(i)}.$$

Lemma 6.22 is proven. □

Now we are ready to prove what is probably the most important property of determinants:

Theorem 6.23. Let $n \in \mathbb{N}$. Let A and B be two $n \times n$-matrices. Then,

$$\det(AB) = \det A \cdot \det B.$$

Proof of Theorem 6.23. Write A and B in the forms $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ and $B = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. The definition of AB thus yields $AB = \left(\sum\limits_{k=1}^{n} a_{i,k} b_{k,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Therefore, (341) (applied to AB and $\sum_{k=1}^{n} a_{i,k} b_{k,j}$ instead of A and $a_{i,j}$) yields

$$\det(AB) = \sum_{\sigma \in S_n} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} \left(\sum_{k=1}^{n} a_{i,k} b_{k,\sigma(i)} \right)}_{\substack{= \sum_{\kappa:[n]\to[n]} \prod_{i=1}^{n} \left(a_{i,\kappa(i)} b_{\kappa(i),\sigma(i)} \right) \\ \text{(by Lemma 6.22, applied to } m=n \\ \text{and } p_{i,k} = a_{i,k} b_{k,\sigma(i)} \text{)}}}$$
$$= \sum_{\sigma \in S_n} (-1)^{\sigma} \sum_{\kappa:[n]\to[n]} \underbrace{\prod_{i=1}^{n} \left(a_{i,\kappa(i)} b_{\kappa(i),\sigma(i)} \right)}_{= \left(\prod_{i=1}^{n} a_{i,\kappa(i)} \right) \left(\prod_{i=1}^{n} b_{\kappa(i),\sigma(i)} \right)}$$
$$= \sum_{\sigma \in S_n} (-1)^{\sigma} \sum_{\kappa:[n]\to[n]} \left(\prod_{i=1}^{n} a_{i,\kappa(i)} \right) \left(\prod_{i=1}^{n} b_{\kappa(i),\sigma(i)} \right)$$
$$= \sum_{\kappa:[n]\to[n]} \left(\prod_{i=1}^{n} a_{i,\kappa(i)} \right) \left(\sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} b_{\kappa(i),\sigma(i)} \right). \tag{367}$$

Now, for every $\kappa : [n] \to [n]$, we let B_κ be the $n \times n$-matrix $\left(b_{\kappa(i),j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Then, for every $\kappa : [n] \to [n]$, the equality (341) (applied to B_κ and $b_{\kappa(i),j}$ instead of A and $a_{i,j}$) yields

$$\det(B_\kappa) = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} b_{\kappa(i),\sigma(i)}. \tag{368}$$

Thus, (367) becomes

$$\det(AB) = \sum_{\kappa:[n]\to[n]} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) \underbrace{\left(\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n b_{\kappa(i),\sigma(i)}\right)}_{\substack{=\det(B_\kappa) \\ \text{(by (368))}}}$$

$$= \sum_{\kappa:[n]\to[n]} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) \det(B_\kappa)$$

$$= \underbrace{\sum_{\substack{\kappa:[n]\to[n]; \\ \kappa\in S_n}}}_{\substack{=\sum_{\kappa\in S_n} \\ \text{(since every } \kappa\in S_n \text{ automatically} \\ \text{is a map } [n]\to[n])}} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) \underbrace{\det(B_\kappa)}_{\substack{=(-1)^\kappa\cdot\det B \\ \text{(by Lemma 6.17 \textbf{(a)})}}}$$

$$+ \sum_{\substack{\kappa:[n]\to[n]; \\ \kappa\notin S_n}} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) \underbrace{\det(B_\kappa)}_{\substack{=0 \\ \text{(by Lemma 6.17 \textbf{(b)})}}}$$

$$= \sum_{\kappa\in S_n} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) (-1)^\kappa \cdot \det B + \underbrace{\sum_{\substack{\kappa:[n]\to[n]; \\ \kappa\notin S_n}} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) 0}_{=0}$$

$$= \sum_{\kappa\in S_n} \left(\prod_{i=1}^n a_{i,\kappa(i)}\right) (-1)^\kappa \cdot \det B = \sum_{\sigma\in S_n} \left(\prod_{i=1}^n a_{i,\sigma(i)}\right) (-1)^\sigma \cdot \det B$$

$$\text{(here, we renamed the summation index } \kappa \text{ as } \sigma\text{)}$$

$$= \underbrace{\left(\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}\right)}_{\substack{=\det A \\ \text{(by (341))}}} \cdot \det B = \det A \cdot \det B.$$

This proves Theorem 6.23. □

Remark 6.24. The analogue of Theorem 6.23 with addition instead of multiplication does not hold. If A and B are two $n\times n$-matrices for some $n\in\mathbb{N}$, then $\det(A+B)$ does usually **not** equal $\det A + \det B$.

We shall now show several applications of Theorem 6.23. First, a simple corollary:

Corollary 6.25. Let $n \in \mathbb{N}$.

(a) If $B_1, B_2, \ldots, B_k$ are finitely many $n \times n$-matrices, then $\det(B_1 B_2 \cdots B_k) = \prod_{i=1}^{k} \det(B_i)$.

(b) If B is any $n \times n$-matrix, and $k \in \mathbb{N}$, then $\det(B^k) = (\det B)^k$.

Proof of Corollary 6.25. Corollary 6.25 easily follows from Theorem 6.23 by induction over k. (The induction base, $k = 0$, relies on the fact that the product of 0 matrices is I_n and has determinant $\det(I_n) = 1$.) We leave the details to the reader. $\square$

Example 6.26. Recall that the Fibonacci sequence is the sequence $(f_0, f_1, f_2, \ldots)$ of integers which is defined recursively by $f_0 = 0$, $f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 2$. We shall prove that

$$f_{n+1} f_{n-1} - f_n^2 = (-1)^n \qquad \text{for every positive integer } n. \tag{369}$$

(This is a classical fact known as the *Cassini identity* and easy to prove by induction, but we shall prove it differently to illustrate the use of determinants.)

Let B be the 2×2-matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ (over the ring $\mathbb{Z}$). It is easy to see that $\det B = -1$. But for every positive integer n, we have

$$B^n = \begin{pmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{pmatrix}. \tag{370}$$

Indeed, (370) can be easily proven by induction over n: For $n = 1$ it is clear by inspection; if it holds for $n = N$, then for $n = N + 1$ it follows from

$$\begin{aligned} B^{N+1} &= \underbrace{B^N}_{\substack{= \begin{pmatrix} f_{N+1} & f_N \\ f_N & f_{N-1} \end{pmatrix} \\ \text{(by the induction hypothesis)}}} \underbrace{B}_{= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}} = \begin{pmatrix} f_{N+1} & f_N \\ f_N & f_{N-1} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} f_{N+1} \cdot 1 + f_N \cdot 1 & f_{N+1} \cdot 1 + f_N \cdot 0 \\ f_N \cdot 1 + f_{N-1} \cdot 1 & f_N \cdot 1 + f_{N-1} \cdot 0 \end{pmatrix} \\ &\qquad \text{(by the definition of a product of two matrices)} \\ &= \begin{pmatrix} f_{N+1} + f_N & f_{N+1} \\ f_N + f_{N-1} & f_N \end{pmatrix} = \begin{pmatrix} f_{N+2} & f_{N+1} \\ f_{N+1} & f_N \end{pmatrix} \end{aligned}$$

(since $f_{N+1} + f_N = f_{N+2}$ and $f_N + f_{N-1} = f_{N+1}$).

Now, let n be a positive integer. Then, (370) yields

$$\det(B^n) = \det \begin{pmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{pmatrix} = f_{n+1} f_{n-1} - f_n^2.$$

On the other hand, Corollary 6.25 **(b)** (applied to $k = n$) yields $\det(B^n) = \left(\underbrace{\det B}_{=-1}\right)^n = (-1)^n$. Hence, $f_{n+1}f_{n-1} - f_n^2 = \det(B^n) = (-1)^n$. This proves (369).

We can generalize (369) as follows:

Exercise 6.10. Let a and b be two complex numbers. Let $(x_0, x_1, x_2, \ldots)$ be a sequence of complex numbers such that every $n \geq 2$ satisfies

$$x_n = ax_{n-1} + bx_{n-2}. \tag{371}$$

(We called such sequences "(a, b)-recurrent" in Definition 4.2.) Let $k \in \mathbb{N}$. Prove that

$$x_{n+1}x_{n-k-1} - x_n x_{n-k} = (-b)^{n-k-1}(x_{k+2}x_0 - x_{k+1}x_1). \tag{372}$$

for every integer $n > k$.

We notice that (369) can be obtained by applying (372) to $a = 1$, $b = 1$, $x_i = f_i$ and $k = 0$. Thus, (372) is a generalization of (369). Notice that you could have easily come up with the identity (372) by trying to generalize the proof of (369) we gave; in contrast, it is not that straightforward to guess the general formula (372) from the classical proof of (369) by induction. So the proof of (369) using determinants has at least the advantage of pointing to a generalization.

Example 6.27. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$, and let $y_1, y_2, \ldots, y_n$ be n further elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $(x_i y_j)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. In Example 6.6, we have shown that $\det A = 0$ if $n \geq 2$. We can now prove this in a simpler way.

Namely, let $n \geq 2$. Define an $n \times n$-matrix B by $B = \begin{pmatrix} x_1 & 0 & 0 & \cdots & 0 \\ x_2 & 0 & 0 & \cdots & 0 \\ x_3 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & 0 & \cdots & 0 \end{pmatrix}$. (Thus, the first column of B is $(x_1, x_2, \ldots, x_n)^T$, while all other columns are filled with zeroes.) Define an $n \times n$-matrix C by $C = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_n \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$. (Thus, the first row of C is $(y_1, y_2, \ldots, y_n)$, while all other rows are filled with zeroes.)

The second row of C consists of zeroes (and this second row indeed exists, because $n \geq 2$). Thus, Exercise 6.7 **(c)** (applied to C instead of A) yields $\det C = 0$.

Similarly, using Exercise 6.7 **(d)**, we can show that $\det B = 0$. Now, Theorem 6.23 (applied to B and C instead of A and B) yields $\det(BC) = \det B \cdot \underbrace{\det C}_{=0} = 0$. But what is BC ?

Write B in the form $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$, and write C in the form $C = (c_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$. Then, the definition of BC yields

$$BC = \left(\sum_{k=1}^{n} b_{i,k}c_{k,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Therefore, for every $(i,j) \in \{1,2,\ldots,n\}^2$, the (i,j)-th entry of the matrix BC is

$$\sum_{k=1}^{n} b_{i,k}c_{k,j} = \underbrace{b_{i,1}}_{=x_i}\underbrace{c_{1,j}}_{=y_j} + \sum_{k=2}^{n}\underbrace{b_{i,k}}_{=0}\underbrace{c_{k,j}}_{=0} = x_iy_j + \underbrace{\sum_{k=2}^{n} 0\cdot 0}_{=0} = x_iy_j.$$

But this is the same as the (i,j)-th entry of the matrix A. Thus, every entry of BC equals the corresponding entry of A. Hence, $BC = A$, so that $\det(BC) = \det A$. Thus, $\det A = \det(BC) = 0$, just as we wanted to show.

Example 6.28. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$, and let $y_1, y_2, \ldots, y_n$ be n further elements of $\mathbb{K}$. Let A be the $n\times n$-matrix $(x_i + y_j)_{1\leq i\leq n,\ 1\leq j\leq n}$. In Example 6.7, we have shown that $\det A = 0$ if $n \geq 3$.

We can now prove this in a simpler way. The argument is similar to Example 6.27, and so I will be very brief:

Let $n \geq 3$. Define an $n\times n$-matrix B by $B = \begin{pmatrix} x_1 & 1 & 0 & \cdots & 0 \\ x_2 & 1 & 0 & \cdots & 0 \\ x_3 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 1 & 0 & \cdots & 0 \end{pmatrix}$. (Thus, the first column of B is $(x_1, x_2, \ldots, x_n)^T$, the second column is $(1,1,\ldots,1)^T$, while all other columns are filled with zeroes.) Define an $n\times n$-matrix C by $C = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ y_1 & y_2 & y_3 & \cdots & y_n \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$. (Thus, the first row of C is $(1,1,\ldots,1)$, the second row is $(y_1, y_2, \ldots, y_n)$, while all other rows are filled with zeroes.) It is now easy to show that $BC = A$ (check this!), but both $\det B$ and $\det C$ are 0 (due to having a column or a row filled with zeroes). Thus, again, we obtain $\det A = 0$.

Exercise 6.11. Let $n \in \mathbb{N}$. Let A be the $n \times n$-matrix $\left(\dbinom{i+j-2}{i-1} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} =$
$$\begin{pmatrix} \dbinom{0}{0} & \dbinom{1}{0} & \cdots & \dbinom{n-1}{0} \\ \dbinom{1}{1} & \dbinom{2}{1} & \cdots & \dbinom{n}{1} \\ \vdots & \vdots & \ddots & \vdots \\ \dbinom{n-1}{n-1} & \dbinom{n}{n-1} & \cdots & \dbinom{2n-2}{n-1} \end{pmatrix}.$$ (This matrix is a piece of Pascal's triangle "rotated by 45°". For example, for $n = 4$, we have $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{pmatrix}$.)

Show that $\det A = 1$.

The matrix A in Exercise 6.11 is one of the so-called *Pascal matrices*; see [EdeStr04] for an enlightening exposition of some of its properties (but beware of the fact that the very first page reveals a significant part of the solution of Exercise 6.11).

Remark 6.29. There exists a more general notion of a matrix, in which the rows and the columns are indexed not necessarily by integers from 1 to n (for some $n \in \mathbb{N}$), but rather by arbitrary objects. For instance, this more general notion allows us to speak of a matrix with two rows labelled "spam" and "eggs", and with three columns labelled 0, 3 and ∞. (It thus has 6 entries, such as the ("spam", 3)-th entry or the ("eggs", ∞)-th entry.) This notion of matrices is more general and more flexible than the one used above (e.g., it allows for infinite matrices), although it has some drawbacks (e.g., notions such as "lower-triangular" are not defined per se, because there might be no canonical way to order the rows and the columns; also, infinite matrices cannot always be multiplied). We might want to define the determinant of such a matrix. Of course, this only makes sense when the rows of the matrix are indexed by the same objects as its columns (this essentially says that the matrix is a "square matrix" in a reasonably general sense). So, let X be a set, and A be a "generalized matrix" whose rows and columns are both indexed by the elements of X. We want to define $\det A$. We assume that X is finite (indeed, while $\det A$ sometimes makes sense for infinite X, this only happens under some rather restrictive conditions). Then, we can define $\det A$ by
$$\det A = \sum_{\sigma \in S_X} (-1)^\sigma \prod_{i \in X} a_{i,\sigma(i)},$$
where S_X denotes the set of all permutations of X. This relies on a definition of $(-1)^\sigma$ for every $\sigma \in S_X$; fortunately, we have provided such a definition in Exercise 5.12.

We shall see more about determinants later. So far we have barely scratched the surface. Huge collections of problems and examples on the computation of determinants can be found in [Prasol94] and [Kratt99] (and, if you can be bothered with 100-years-old notation and level of rigor, in Muir's five-volume book series [Muir30] – one of the most comprehensive collections of "forgotten tales" in mathematics[224]).

Let us finish this section with a brief remark on the geometrical use of determinants.

Remark 6.30. Let us consider the Euclidean plane $\mathbb{R}^2$ with its Cartesian coordinate system and its origin 0. If $A = (x_A, y_A)$ and $B = (x_B, y_B)$ are two points on $\mathbb{R}^2$, then the area of the triangle $0AB$ is $\dfrac{1}{2}\left|\det\begin{pmatrix} x_A & x_B \\ y_A & y_B \end{pmatrix}\right|$. The absolute value here reflects the fact that determinants can be negative, while areas must always be ≥ 0 (although they can be 0 when 0, A and B are collinear); however, it makes working with areas somewhat awkward. This can be circumvented by the notion of a *signed area*. (The signed area of a triangle ABC is its regular area if the triangle is "directed clockwise", and otherwise it is the negative of its area.) The signed area of the triangle $0AB$ is $\dfrac{1}{2}\det\begin{pmatrix} x_A & x_B \\ y_A & y_B \end{pmatrix}$. (Here, 0 stands for the origin; i.e., "the triangle $0AB$" means the triangle with vertices at the origin, at A and at B.)

If $A = (x_A, y_A)$, $B = (x_B, y_B)$ and $C = (x_C, y_C)$ are three points in $\mathbb{R}^2$, then the signed area of triangle ABC is $\dfrac{1}{2}\det\begin{pmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{pmatrix}$.

Similar formulas hold for tetrahedra: If $A = (x_A, y_A, z_A)$, $B = (x_B, y_B, z_B)$ and $C = (x_C, y_C, z_C)$ are three points in $\mathbb{R}^3$, then the signed volume of the tetrahedron $0ABC$ is $\dfrac{1}{6}\det\begin{pmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ z_A & z_B & z_C \end{pmatrix}$. (Again, take the absolute value for the non-signed volume.) There is a 4×4 determinant formula for the signed volume of a general tetrahedron $ABCD$.

One can generalize the notion of a triangle in $\mathbb{R}^2$ and the notion of a tetrahedron in $\mathbb{R}^3$ to a notion of a *simplex* in $\mathbb{R}^n$. Then, one can try to define a notion of volume for these objects. Determinants provide a way to do this. (Obviously,

[224] In this series, Muir endeavors to summarize every paper that had been written about determinants until the year 1920. Several of these papers contain results that have fallen into oblivion, and not always justly so; Muir's summaries are thus a goldmine of interesting material. However, his notation is antiquated and his exposition is often extremely unintelligible (e.g., complicated identities are often presented by showing an example and hoping that the reader will correctly guess the pattern); very few proofs are given.

Three other classical British texts on determinants are Muir's and Metzler's [MuiMet60], Turnbull's [Turnbu29] and Aitken's [Aitken56]; these texts (particularly the first two) contain a wealth of remarkable results, many of which are barely remembered today. Unfortunately, their clarity and their level of rigor leave much to be desired by modern standards.

they don't allow you to define the volume of a general "convex body" like a sphere, and even for simplices it is not a-priori clear that they satisfy the standard properties that one would expect them to have – e.g., that the "volume" of a simplex does not change when one moves this simplex. But for the algebraic part of analytic geometry, they are mostly sufficient. To define "volumes" for general convex bodies, one needs calculus and the theory of integration in $\mathbb{R}^n$; but this theory, too, uses determinants.)

6.5. The Cauchy-Binet formula

This section is devoted to the Cauchy-Binet formula: a generalization of Theorem 6.23 which is less well-known than the latter, but still comes useful. This formula appears in the literature in various forms; we follow the one on PlanetMath (although we use different notations).

First, we introduce a notation for "picking out some rows of a matrix and throwing away the rest" (and also the analogous thing for columns):

Definition 6.31. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$ be an $n \times m$-matrix.

(a) If $i_1, i_2, \ldots, i_u$ are some elements of $\{1, 2, \ldots, n\}$, then we let $\operatorname{rows}_{i_1, i_2, \ldots, i_u} A$ denote the $u \times m$-matrix $\left(a_{i_x,j}\right)_{1 \leq x \leq u,\ 1 \leq j \leq m}$. For instance, if $A = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix}$, then $\operatorname{rows}_{3,1,4} A = \begin{pmatrix} c & c' & c'' \\ a & a' & a'' \\ d & d' & d'' \end{pmatrix}$. For every $p \in \{1, 2, \ldots, u\}$, we have

$$\begin{aligned} & \left(\text{the } p\text{-th row of } \operatorname{rows}_{i_1, i_2, \ldots, i_u} A\right) \\ &= \left(a_{i_p,1}, a_{i_p,2}, \ldots, a_{i_p,m}\right) && \left(\text{since } \operatorname{rows}_{i_1, i_2, \ldots, i_u} A = \left(a_{i_x,j}\right)_{1 \leq x \leq u,\ 1 \leq j \leq m}\right) \\ &= \left(\text{the } i_p\text{-th row of } A\right) && \left(\text{since } A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}\right). \end{aligned} \tag{373}$$

Thus, $\operatorname{rows}_{i_1, i_2, \ldots, i_u} A$ is the $u \times m$-matrix whose rows (from top to bottom) are the rows labelled $i_1, i_2, \ldots, i_u$ of the matrix A.

(b) If $j_1, j_2, \ldots, j_v$ are some elements of $\{1, 2, \ldots, m\}$, then we let $\operatorname{cols}_{j_1, j_2, \ldots, j_v} A$ denote the $n \times v$-matrix $\left(a_{i,j_y}\right)_{1 \leq i \leq n,\ 1 \leq y \leq v}$. For instance, if $A = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix}$, then $\operatorname{cols}_{3,2} A = \begin{pmatrix} a'' & a' \\ b'' & b' \\ c'' & c' \end{pmatrix}$. For every $q \in \{1, 2, \ldots, v\}$, we have

$$\begin{aligned} &\left(\text{the } q\text{-th column of } \operatorname{cols}_{j_1, j_2, \ldots, j_v} A\right) \\ &= \begin{pmatrix} a_{1,j_q} \\ a_{2,j_q} \\ \vdots \\ a_{n,j_q} \end{pmatrix} \qquad \left(\text{since } \operatorname{cols}_{j_1, j_2, \ldots, j_v} A = \left(a_{i,j_y}\right)_{1 \leq i \leq n,\ 1 \leq y \leq v}\right) \\ &= \left(\text{the } j_q\text{-th column of } A\right) \qquad \left(\text{since } A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}\right). \end{aligned} \tag{374}$$

Thus, $\operatorname{cols}_{j_1, j_2, \ldots, j_v} A$ is the $n \times v$-matrix whose columns (from left to right) are the columns labelled $j_1, j_2, \ldots, j_v$ of the matrix A.

Now we can state the *Cauchy-Binet formula*:

Theorem 6.32. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix, and let B be an $m \times n$-matrix. Then,

$$\det(AB) = \sum_{1 \leq g_1 < g_2 < \cdots < g_n \leq m} \det\left(\operatorname{cols}_{g_1, g_2, \ldots, g_n} A\right) \cdot \det\left(\operatorname{rows}_{g_1, g_2, \ldots, g_n} B\right). \tag{375}$$

Remark 6.33. The summation sign $\sum\limits_{1 \leq g_1 < g_2 < \cdots < g_n \leq m}$ in (375) is an abbreviation for

$$\sum_{\substack{(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, m\}^n; \\ g_1 < g_2 < \cdots < g_n}}. \tag{376}$$

In particular, if $n = 0$, then it signifies a summation over all 0-tuples of elements of $\{1, 2, \ldots, m\}$ (because in this case, the chain of inequalities $g_1 < g_2 < \cdots < g_n$ is a tautology); such a sum always has exactly one addend (because there is exactly one 0-tuple).

When both n and m equal 0, then the notation $\sum\limits_{1 \leq g_1 < g_2 < \cdots < g_n \leq m}$ is slightly confusing: It appears to mean an empty summation (because $1 \leq m$ does not hold). But as we said, we mean this notation to be an abbreviation for (376), which signifies a sum with exactly one addend. But this is enough pedantry for now; for $n > 0$, the notation $\sum\limits_{1 \leq g_1 < g_2 < \cdots < g_n \leq m}$ fortunately means exactly what it seems to mean.

We shall soon give a detailed proof of Theorem 6.32; see [AigZie14, Chapter 32, Theorem] for a different proof[225]. Before we prove Theorem 6.32, let us give some examples for its use. First, here is a simple fact:

Lemma 6.34. Let $n \in \mathbb{N}$.
(a) There exists exactly one n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$, namely the n-tuple $(1, 2, \ldots, n)$.
(b) Let $m \in \mathbb{N}$ be such that $m < n$. Then, there exists no n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, m\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$.

As for its intuitive meaning, Lemma 6.34 can be viewed as a "pigeonhole principle" for strictly increasing sequences: Part **(b)** says (roughly speaking) that there is no way to squeeze a strictly increasing sequence $(g_1, g_2, \ldots, g_n)$ of n numbers into the set $\{1, 2, \ldots, m\}$ when $m < n$; part **(a)** says (again, informally) that the only such sequence for $m = n$ is $(1, 2, \ldots, n)$.

Exercise 6.12. Give a formal proof of Lemma 6.34. (Do not bother doing this if you do not particularly care about formal proofs and find Lemma 6.34 obvious enough.)

Example 6.35. Let $n \in \mathbb{N}$. Let A and B be two $n \times n$-matrices. It is easy to check that $\operatorname{cols}_{1,2,\ldots,n} A = A$ and $\operatorname{rows}_{1,2,\ldots,n} B = B$. Now, Theorem 6.32 (applied to $m = n$) yields

$$\det(AB) = \sum_{1 \leq g_1 < g_2 < \cdots < g_n \leq n} \det\left(\operatorname{cols}_{g_1, g_2, \ldots, g_n} A\right) \cdot \det\left(\operatorname{rows}_{g_1, g_2, \ldots, g_n} B\right). \qquad (378)$$

But Lemma 6.34 **(a)** yields that there exists exactly one n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$, namely the n-tuple $(1, 2, \ldots, n)$. Hence, the sum on the right hand side of (378) has exactly one addend: namely, the addend for $(g_1, g_2, \ldots, g_n) = (1, 2, \ldots, n)$. Therefore, this sum simplifies as

[225] Note that the formulation of Theorem 6.32 in [AigZie14, Chapter 32, Theorem] is slightly different: In our notations, it says that if A is an $n \times m$-matrix and if B is an $m \times n$-matrix, then

$$\det(AB) = \sum_{\substack{\mathcal{Z} \subseteq \{1, 2, \ldots, m\}; \\ |\mathcal{Z}| = n}} \det(\operatorname{cols}_{\mathcal{Z}} A) \cdot \det(\operatorname{rows}_{\mathcal{Z}} B), \qquad (377)$$

where the matrices $\operatorname{cols}_{\mathcal{Z}} A$ and $\operatorname{rows}_{\mathcal{Z}} B$ (for $\mathcal{Z}$ being a subset of $\{1, 2, \ldots, m\}$) are defined as follows: Write the subset $\mathcal{Z}$ in the form $\{z_1, z_2, \ldots, z_k\}$ with $z_1 < z_2 < \cdots < z_k$, and set $\operatorname{cols}_{\mathcal{Z}} A = \operatorname{cols}_{z_1, z_2, \ldots, z_k} A$ and $\operatorname{rows}_{\mathcal{Z}} B = \operatorname{rows}_{z_1, z_2, \ldots, z_k} B$. (Apart from this, [AigZie14, Chapter 32, Theorem] also requires $n \leq m$; but this requirement is useless.)

The equalities (375) and (377) are equivalent, because the n-tuples $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, m\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ are in a bijection with the subsets $\mathcal{Z}$ of $\{1, 2, \ldots, m\}$ satisfying $|\mathcal{Z}| = n$. (This bijection sends an n-tuple $(g_1, g_2, \ldots, g_n)$ to the subset $\{g_1, g_2, \ldots, g_n\}$.)

The proof of (377) given in [AigZie14, Chapter 32] uses the *Lindström-Gessel-Viennot lemma* (which it calls the "lemma of Gessel-Viennot") and is highly worth reading.

follows:

$$\sum_{1\leq g_1<g_2<\cdots<g_n\leq n} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right)\cdot\det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right)$$
$$=\det\left(\underbrace{\operatorname{cols}_{1,2,\ldots,n} A}_{=A}\right)\cdot\det\left(\underbrace{\operatorname{rows}_{1,2,\ldots,n} B}_{=B}\right)=\det A\cdot\det B.$$

Hence, (378) becomes

$$\det(AB)=\sum_{1\leq g_1<g_2<\cdots<g_n\leq n} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right)\cdot\det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right)$$
$$=\det A\cdot\det B.$$

This, of course, is the statement of Theorem 6.23. Hence, Theorem 6.23 is a particular case of Theorem 6.32.

Example 6.36. Let $n\in\mathbb{N}$ and $m\in\mathbb{N}$ be such that $m<n$. Thus, Lemma 6.34 **(b)** shows that there exists no n-tuple $(g_1,g_2,\ldots,g_n)\in\{1,2,\ldots,m\}^n$ satisfying $g_1<g_2<\cdots<g_n$.

Now, let A be an $n\times m$-matrix, and let B be an $m\times n$-matrix. Then, Theorem 6.32 yields

$$\det(AB)=\sum_{1\leq g_1<g_2<\cdots<g_n\leq m} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right)\cdot\det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right)$$
$$=(\text{empty sum})\qquad\left(\begin{array}{c}\text{since there exists no } n\text{-tuple } (g_1,g_2,\ldots,g_n)\in\{1,2,\ldots,m\}^n\\ \text{satisfying } g_1<g_2<\cdots<g_n\end{array}\right)$$
$$=0. \tag{379}$$

This identity allows us to compute $\det A$ in Example 6.27 in a simpler way: Instead of defining two $n\times n$-matrices B and C by $B=\begin{pmatrix} x_1 & 0 & 0 & \cdots & 0\\ x_2 & 0 & 0 & \cdots & 0\\ x_3 & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ x_n & 0 & 0 & \cdots & 0\end{pmatrix}$ and $C=\begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_n\\ 0 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 0\end{pmatrix}$, it suffices to define an $n\times 1$-matrix B' by $B'=(x_1,x_2,\ldots,x_n)^T$ and a $1\times n$-matrix C' by $C'=(y_1,y_2,\ldots,y_n)$, and argue that $A = B'C'$. (We leave the details to the reader.) Similarly, Example 6.28 could be dealt with.

Remark 6.37. The equality (379) can also be derived from Theorem 6.23. Indeed, let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that $m < n$. Let A be an $n \times m$-matrix, and let B be an $m \times n$-matrix. Notice that $n - m > 0$ (since $m < n$). Let A' be the $n \times n$-matrix obtained from A by appending $n - m$ new columns to the right of A and filling these columns with zeroes. (For example, if $n = 4$ and $m = 2$ and $A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \\ a_{4,1} & a_{4,2} \end{pmatrix}$, then $A' = \begin{pmatrix} a_{1,1} & a_{1,2} & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 \\ a_{3,1} & a_{3,2} & 0 & 0 \\ a_{4,1} & a_{4,2} & 0 & 0 \end{pmatrix}$.) Also, let B' be the $n \times n$-matrix obtained from B by appending $n - m$ new rows to the bottom of B and filling these rows with zeroes. (For example, if $n = 4$ and $m = 2$ and $B = \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} \end{pmatrix}$, then $B' = \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.) Then, it is easy to check that $AB = A'B'$ (in fact, just compare corresponding entries of AB and $A'B'$). But recall that $n - m > 0$. Hence, the matrix A' has a column consisting of zeroes (namely, its last column). Thus, Exercise 6.7 **(d)** (applied to A' instead of A) shows that $\det(A') = 0$. Now,

$$\det\left(\underbrace{AB}_{=A'B'}\right) = \det(A'B') = \underbrace{\det(A')}_{=0} \cdot \det(B')$$
$$\left(\text{by Theorem 6.23, applied to } A' \text{ and } B' \text{ instead of } A \text{ and } B\right)$$
$$= 0.$$

Thus, (379) is proven again.

Example 6.38. Let us see what Theorem 6.32 says for $n = 1$. Indeed, let $m \in \mathbb{N}$; let $A = (a_1, a_2, \ldots, a_m)$ be a $1 \times m$-matrix (i.e., a row vector with m entries), and let $B = (b_1, b_2, \ldots, b_m)^T$ be an $m \times 1$-matrix (i.e., a column vector with m entries). Then, AB is the 1×1-matrix $\left(\sum\limits_{k=1}^{m} a_k b_k\right)$. Thus,

$$\det(AB) = \det\left(\sum_{k=1}^{m} a_k b_k\right) = \sum_{k=1}^{m} a_k b_k \qquad (\text{by } (342)). \tag{380}$$

What would we obtain if we tried to compute $\det(AB)$ using Theorem 6.32?

Theorem 6.32 (applied to $n = 1$) yields

$$\det(AB) = \underbrace{\sum_{1 \leq g_1 \leq m}}_{= \sum_{g_1=1}^{m}} \det\left(\underbrace{\operatorname{cols}_{g_1} A}_{=\left(\begin{array}{c} a_{g_1} \end{array}\right)}\right) \cdot \det\left(\underbrace{\operatorname{rows}_{g_1} B}_{=\left(\begin{array}{c} b_{g_1} \end{array}\right)}\right)$$
$$= \sum_{g_1=1}^{m} \underbrace{\det\left(\begin{array}{c} a_{g_1} \end{array}\right)}_{\substack{=a_{g_1} \\ \text{(by (342))}}} \cdot \underbrace{\det\left(\begin{array}{c} b_{g_1} \end{array}\right)}_{\substack{=b_{g_1} \\ \text{(by (342))}}} = \sum_{g_1=1}^{m} a_{g_1} \cdot b_{g_1}.$$

This is, of course, the same result as that of (380) (with the summation index k renamed as g_1). So we did not gain any interesting insight from applying Theorem 6.32 to $n = 1$.

Example 6.39. Let us try a slightly less trivial case. Indeed, let $m \in \mathbb{N}$; let $A = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ a'_1 & a'_2 & \cdots & a'_m \end{pmatrix}$ be a $2 \times m$-matrix, and let $B = \begin{pmatrix} b_1 & b'_1 \\ b_2 & b'_2 \\ \vdots & \vdots \\ b_m & b'_m \end{pmatrix}$ be an $m \times 2$-matrix. Then, AB is the 2×2-matrix $\begin{pmatrix} \sum_{k=1}^{m} a_k b_k & \sum_{k=1}^{m} a_k b'_k \\ \sum_{k=1}^{m} a'_k b_k & \sum_{k=1}^{m} a'_k b'_k \end{pmatrix}$. Hence,

$$\det(AB) = \det \begin{pmatrix} \sum_{k=1}^{m} a_k b_k & \sum_{k=1}^{m} a_k b'_k \\ \sum_{k=1}^{m} a'_k b_k & \sum_{k=1}^{m} a'_k b'_k \end{pmatrix}$$
$$= \left(\sum_{k=1}^{m} a_k b_k\right) \left(\sum_{k=1}^{m} a'_k b'_k\right) - \left(\sum_{k=1}^{m} a'_k b_k\right) \left(\sum_{k=1}^{m} a_k b'_k\right). \qquad (381)$$

On the other hand, Theorem 6.32 (now applied to $n = 2$) yields

$$\det(AB) = \sum_{1 \leq g_1 < g_2 \leq m} \det\left(\operatorname{cols}_{g_1,g_2} A\right) \cdot \det\left(\operatorname{rows}_{g_1,g_2} B\right)$$

$$= \sum_{1 \leq i < j \leq m} \det\left(\underbrace{\operatorname{cols}_{i,j} A}_{=\begin{pmatrix} a_i & a_j \\ a'_i & a'_j \end{pmatrix}}\right) \cdot \det\left(\underbrace{\operatorname{rows}_{i,j} B}_{=\begin{pmatrix} b_i & b'_i \\ b_j & b'_j \end{pmatrix}}\right)$$

$$\left(\begin{array}{c}\text{here, we renamed the summation indices } g_1 \text{ and } g_2 \\ \text{as } i \text{ and } j\text{, since double subscripts are annoying}\end{array}\right)$$

$$= \sum_{1 \leq i < j \leq m} \underbrace{\det\begin{pmatrix} a_i & a_j \\ a'_i & a'_j \end{pmatrix}}_{=a_i a'_j - a_j a'_i} \cdot \underbrace{\det\begin{pmatrix} b_i & b'_i \\ b_j & b'_j \end{pmatrix}}_{=b_i b'_j - b_j b'_i}$$

$$= \sum_{1 \leq i < j \leq m} \left(a_i a'_j - a_j a'_i\right) \cdot \left(b_i b'_j - b_j b'_i\right).$$

Compared with (381), this yields

$$\left(\sum_{k=1}^{m} a_k b_k\right)\left(\sum_{k=1}^{m} a'_k b'_k\right) - \left(\sum_{k=1}^{m} a'_k b_k\right)\left(\sum_{k=1}^{m} a_k b'_k\right) = \sum_{1 \leq i < j \leq m} \left(a_i a'_j - a_j a'_i\right) \cdot \left(b_i b'_j - b_j b'_i\right). \tag{382}$$

This identity is called the *Binet-Cauchy identity* (I am not kidding – look it up on the Wikipedia). It is fairly easy to prove by direct computation; thus, using Theorem 6.32 to prove it was quite an overkill. However, (382) might not be very easy to come up with, whereas deriving it from Theorem 6.32 is straightforward. (And Theorem 6.32 is easier to memorize than (382).)

Here is a neat application of (382): If $a_1, a_2, \ldots, a_m$ and $a'_1, a'_2, \ldots, a'_m$ are real numbers, then (382) (applied to $b_k = a_k$ and $b'_k = a'_k$) yields

$$\left(\sum_{k=1}^{m} a_k a_k\right)\left(\sum_{k=1}^{m} a'_k a'_k\right) - \left(\sum_{k=1}^{m} a'_k a_k\right)\left(\sum_{k=1}^{m} a_k a'_k\right) = \sum_{1 \leq i < j \leq m} \underbrace{\left(a_i a'_j - a_j a'_i\right) \cdot \left(a_i a'_j - a_j a'_i\right)}_{=\left(a_i a'_j - a_j a'_i\right)^2 \geq 0} \geq \sum_{1 \leq i < j \leq m} 0 = 0,$$

so that

$$\left(\sum_{k=1}^{m} a_k a_k\right)\left(\sum_{k=1}^{m} a'_k a'_k\right) \geq \left(\sum_{k=1}^{m} a'_k a_k\right)\left(\sum_{k=1}^{m} a_k a'_k\right).$$

In other words,

$$\left(\sum_{k=1}^{m} a_k^2\right)\left(\sum_{k=1}^{m} \left(a_k'\right)^2\right) \geq \left(\sum_{k=1}^{m} a_k a_k'\right)^2 .$$

This is the famous Cauchy-Schwarz inequality.

Let us now prepare for the proof of Theorem 6.32. First comes a fact which should be fairly clear:

Proposition 6.40. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n integers.

(a) There exists a permutation $\sigma \in S_n$ such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$.

(b) If $\sigma \in S_n$ is such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$, then, for every $i \in \{1, 2, \ldots, n\}$, the value $a_{\sigma(i)}$ depends only on $a_1, a_2, \ldots, a_n$ and i (but not on σ).

(c) Assume that the integers $a_1, a_2, \ldots, a_n$ are distinct. Then, there is a **unique** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$.

Let me explain why this proposition should be intuitively obvious.[226] Proposition 6.40 **(a)** says that any list $(a_1, a_2, \ldots, a_n)$ of n integers can be sorted in weakly increasing order by means of a permutation $\sigma \in S_n$. Proposition 6.40 **(b)** says that the result of this sorting process is independent of how the sorting happened (although the permutation σ will sometimes be non-unique). Proposition 6.40 **(c)** says that if the integers $a_1, a_2, \ldots, a_n$ are distinct, then the permutation $\sigma \in S_n$ which sorts the list $(a_1, a_2, \ldots, a_n)$ in increasing order is uniquely determined as well. We required $a_1, a_2, \ldots, a_n$ to be n integers for the sake of simplicity, but we could just as well have required them to be elements of any *totally ordered set* (i.e., any set with a less-than relation satisfying some standard axioms).

The next fact looks slightly scary, but is still rather simple:

Lemma 6.41. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. We let $\mathbf{E}$ be the subset

$$\left\{(k_1, k_2, \ldots, k_n) \in [m]^n \mid \text{ the integers } k_1, k_2, \ldots, k_n \text{ are distinct}\right\}$$

of $[m]^n$. We let $\mathbf{I}$ be the subset

$$\left\{(k_1, k_2, \ldots, k_n) \in [m]^n \mid k_1 < k_2 < \cdots < k_n\right\}$$

of $[m]^n$. Then, the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

is well-defined and is a bijection.

[226] See the solution of Exercise 6.13 further below for a formal proof of this proposition.

The intuition for Lemma 6.41 is that every n-tuple of distinct elements of $\{1, 2, \ldots, m\}$ can be represented uniquely as a permuted version of a strictly increasing[227] n-tuple of elements of $\{1, 2, \ldots, m\}$, and therefore, specifying an n-tuple of distinct elements of $\{1, 2, \ldots, m\}$ is tantamount to specifying a strictly increasing n-tuple of elements of $\{1, 2, \ldots, m\}$ and a permutation $\sigma \in S_n$ which says how this n-tuple is to be permuted.[228] This is not a formal proof, but this should explain why Lemma 6.41 is usually applied throughout mathematics without even mentioning it as a statement. If desired, a formal proof of Lemma 6.41 can be obtained using Proposition 6.40.[229]

Exercise 6.13. Prove Proposition 6.40 and Lemma 6.41. (Ignore this exercise if you find these two facts sufficiently obvious and are uninterested in the details of their proofs.)

Before we return to Theorem 6.32, let me make a digression about sorting:

Exercise 6.14. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that $n \geq m$. Let $a_1, a_2, \ldots, a_n$ be n integers. Let $b_1, b_2, \ldots, b_m$ be m integers. Assume that

$$a_i \leq b_i \qquad \text{for every } i \in \{1, 2, \ldots, m\}. \tag{383}$$

Let $\sigma \in S_n$ be such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. Let $\tau \in S_m$ be such that $b_{\tau(1)} \leq b_{\tau(2)} \leq \cdots \leq b_{\tau(m)}$. Then,

$$a_{\sigma(i)} \leq b_{\tau(i)} \qquad \text{for every } i \in \{1, 2, \ldots, m\}.$$

Remark 6.42. Loosely speaking, Exercise 6.14 says the following: If two lists $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_m)$ of integers have the property that each entry of the first list is $\leq$ to the corresponding entry of the second list (as long as the latter is well-defined), then this property still holds after both lists are sorted in increasing order, provided that we have $n \geq m$ (that is, the first list is at least as long as the second list).

A consequence of Exercise 6.14 is the following curious fact, known as the "non-messing-up phenomenon" ([Tenner04, Theorem 1] and [GalKar71, Example 1]): If we start with a matrix filled with integers, then sort the entries of each

[227] An n-tuple $(k_1, k_2, \ldots, k_n)$ is said to be *strictly increasing* if and only if $k_1 < k_2 < \cdots < k_n$.

[228] For instance, the 4-tuple $(4, 1, 6, 2)$ of distinct elements of $\{1, 2, \ldots, 7\}$ can be specified by specifying the strictly increasing 4-tuple $(1, 2, 4, 6)$ (which is its sorted version) and the permutation $\pi \in S_4$ which sends $1, 2, 3, 4$ to $3, 1, 4, 2$, respectively (that is, $\pi = (3, 1, 4, 2)$ in one-line notation). In the terminology of Lemma 6.41, the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

sends $((1, 2, 4, 6), \pi)$ to $(4, 1, 6, 2)$.

[229] Again, see the solution of Exercise 6.13 further below for such a proof.

row of the matrix in increasing order, and then sort the entries of each column of the resulting matrix in increasing order, then the final matrix still has sorted rows (i.e., the entries of each row are still sorted). That is, the sorting of the columns did not "mess up" the sortedness of the rows. For example, if we start with the matrix $\begin{pmatrix} 1 & 3 & 2 & 5 \\ 2 & 1 & 4 & 2 \\ 3 & 1 & 6 & 0 \end{pmatrix}$, then sorting the entries of each row gives us the matrix $\begin{pmatrix} 1 & 2 & 3 & 5 \\ 1 & 2 & 2 & 4 \\ 0 & 1 & 3 & 6 \end{pmatrix}$, and then sorting the entries of each column results in the matrix $\begin{pmatrix} 0 & 1 & 2 & 4 \\ 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 6 \end{pmatrix}$. The rows of this matrix are still sorted, as the "non-messing-up phenomenon" predicts. To prove this phenomenon in general, it suffices to show that any entry in the resulting matrix is $\leq$ to the entry directly below it (assuming that the latter exists); but this follows easily from Exercise 6.14.

We are now ready to prove Theorem 6.32.

Proof of Theorem 6.32. We shall use the notations of Lemma 6.41.

Write the $n \times m$-matrix A as $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Write the $m \times n$-matrix B as $B = \left(b_{i,j}\right)_{1 \leq i \leq m,\ 1 \leq j \leq n}$. The definition of AB thus yields $AB = \left(\sum\limits_{k=1}^{m} a_{i,k} b_{k,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Therefore, (341) (applied to AB and $\sum\limits_{k=1}^{m} a_{i,k} b_{k,j}$ instead of A and $a_{i,j}$) yields

$$\det(AB) = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} \left(\sum_{k=1}^{m} a_{i,k} b_{k,\sigma(i)} \right). \tag{384}$$

But for every $\sigma \in S_n$, we have

$$\begin{aligned}
\prod_{i=1}^{n} \left(\sum_{k=1}^{m} a_{i,k} b_{k,\sigma(i)} \right) &= \sum_{(k_1,k_2,\ldots,k_n) \in [m]^n} \underbrace{\prod_{i=1}^{n} \left(a_{i,k_i} b_{k_i,\sigma(i)} \right)}_{= \left(\prod_{i=1}^{n} a_{i,k_i}\right) \left(\prod_{i=1}^{n} b_{k_i,\sigma(i)}\right)} \\
&\qquad \left(\text{by Lemma 6.20, applied to } m_i = n \text{ and } p_{i,k} = a_{i,k} b_{k,\sigma(i)}\right) \\
&= \sum_{(k_1,k_2,\ldots,k_n) \in [m]^n} \left(\prod_{i=1}^{n} a_{i,k_i} \right) \left(\prod_{i=1}^{n} b_{k_i,\sigma(i)} \right).
\end{aligned}$$

Hence, (384) rewrites as

$$\det(AB) = \sum_{\sigma\in S_n} (-1)^\sigma \sum_{(k_1,k_2,\ldots,k_n)\in[m]^n} \left(\prod_{i=1}^n a_{i,k_i}\right)\left(\prod_{i=1}^n b_{k_i,\sigma(i)}\right)$$
$$= \underbrace{\sum_{\sigma\in S_n} \sum_{(k_1,k_2,\ldots,k_n)\in[m]^n}}_{=\sum_{(k_1,k_2,\ldots,k_n)\in[m]^n}\sum_{\sigma\in S_n}} (-1)^\sigma \left(\prod_{i=1}^n a_{i,k_i}\right)\left(\prod_{i=1}^n b_{k_i,\sigma(i)}\right)$$
$$= \sum_{(k_1,k_2,\ldots,k_n)\in[m]^n} \sum_{\sigma\in S_n} (-1)^\sigma \left(\prod_{i=1}^n a_{i,k_i}\right)\left(\prod_{i=1}^n b_{k_i,\sigma(i)}\right)$$
$$= \sum_{(k_1,k_2,\ldots,k_n)\in[m]^n} \left(\prod_{i=1}^n a_{i,k_i}\right)\left(\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n b_{k_i,\sigma(i)}\right). \tag{385}$$

But every $(k_1,k_2,\ldots,k_n)\in[m]^n$ satisfies

$$\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n b_{k_i,\sigma(i)} = \det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) \tag{386}$$

[230]. Hence, (385) becomes

$$\det(AB) = \sum_{(k_1,k_2,\ldots,k_n)\in[m]^n} \left(\prod_{i=1}^n a_{i,k_i}\right)\underbrace{\left(\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n b_{k_i,\sigma(i)}\right)}_{=\det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right)}$$
$$= \sum_{(k_1,k_2,\ldots,k_n)\in[m]^n} \left(\prod_{i=1}^n a_{i,k_i}\right)\det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right). \tag{387}$$

But for every $(k_1,k_2,\ldots,k_n)\in[m]^n$ satisfying $(k_1,k_2,\ldots,k_n)\notin\mathbf{E}$, we have

$$\det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) = 0 \tag{388}$$

[230] *Proof.* Let $(k_1,k_2,\ldots,k_n)\in[m]^n$. Recall that $B = \left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq n}$. Hence, the definition of $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$ gives us

$$\operatorname{rows}_{k_1,k_2,\ldots,k_n} B = \left(b_{k_x,j}\right)_{1\leq x\leq n,\ 1\leq j\leq n} = \left(b_{k_i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$$

(here, we renamed the index x as i). Hence, (341) (applied to $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$ and $b_{k_i,j}$ instead of A and $a_{i,j}$) yields

$$\det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n b_{k_i,\sigma(i)},$$

qed.

[231]. Therefore, in the sum on the right hand side of (387), all the addends corresponding to $(k_1, k_2, \ldots, k_n) \in [m]^n$ satisfying $(k_1, k_2, \ldots, k_n) \notin \mathbf{E}$ evaluate to 0. We can therefore remove all these addends from the sum. The remaining addends are those corresponding to $(k_1, k_2, \ldots, k_n) \in \mathbf{E}$. Therefore, (387) becomes

$$\det(AB) = \sum_{(k_1,k_2,\ldots,k_n)\in\mathbf{E}} \left(\prod_{i=1}^{n} a_{i,k_i}\right) \det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right). \tag{389}$$

On the other hand, Lemma 6.41 yields that the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

is well-defined and is a bijection. Hence, we can substitute $\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$ for $(k_1, k_2, \ldots, k_n)$ in the sum on the right hand side of (389). We thus obtain

$$\begin{aligned}
&\sum_{(k_1,k_2,\ldots,k_n)\in\mathbf{E}} \left(\prod_{i=1}^{n} a_{i,k_i}\right) \det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) \\
&= \sum_{((g_1,g_2,\ldots,g_n),\sigma)\in\mathbf{I}\times S_n} \left(\prod_{i=1}^{n} a_{i,g_{\sigma(i)}}\right) \det\left(\operatorname{rows}_{g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}} B\right).
\end{aligned}$$

Thus, (389) becomes

$$\begin{aligned}
\det(AB) &= \sum_{(k_1,k_2,\ldots,k_n)\in\mathbf{E}} \left(\prod_{i=1}^{n} a_{i,k_i}\right) \det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) \\
&= \sum_{((g_1,g_2,\ldots,g_n),\sigma)\in\mathbf{I}\times S_n} \left(\prod_{i=1}^{n} a_{i,g_{\sigma(i)}}\right) \det\left(\operatorname{rows}_{g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}} B\right).
\end{aligned} \tag{390}$$

But every $(k_1, k_2, \ldots, k_n) \in [m]^n$ and every $\sigma \in S_n$ satisfy

$$\det\left(\operatorname{rows}_{k_{\sigma(1)},k_{\sigma(2)},\ldots,k_{\sigma(n)}} B\right) = (-1)^{\sigma} \cdot \det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) \tag{391}$$

[231] *Proof of (388):* Let $(k_1, k_2, \ldots, k_n) \in [m]^n$ be such that $(k_1, k_2, \ldots, k_n) \notin \mathbf{E}$. Then, the integers $k_1, k_2, \ldots, k_n$ are not distinct (because $\mathbf{E}$ is the set of all n-tuples in $[m]^n$ whose entries are distinct). Thus, there exist two distinct elements p and q of $[n]$ such that $k_p = k_q$. Consider these p and q. But $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$ is the $n \times n$-matrix whose rows (from top to bottom) are the rows labelled $k_1, k_2, \ldots, k_n$ of the matrix B. Since $k_p = k_q$, this shows that the p-th row and the q-th row of the matrix $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$ are equal. Hence, the matrix $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$ has two equal rows (since p and q are distinct). Therefore, Exercise 6.7 **(e)** (applied to $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$ instead of A) yields $\det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right) = 0$, qed.

[232]. Hence, (390) becomes

$$\begin{aligned}
&\det(AB) \\
&= \underbrace{\sum_{((g_1,g_2,\ldots,g_n),\sigma)\in \mathbf{I}\times S_n}}_{=\sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}}\sum_{\sigma\in S_n}} \left(\prod_{i=1}^n a_{i,g_{\sigma(i)}}\right) \underbrace{\det\left(\operatorname{rows}_{g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}} B\right)}_{\substack{=(-1)^\sigma\cdot\det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right) \\ \text{(by (391), applied to } k_i=g_i\text{)}}} \\
&= \sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}} \sum_{\sigma\in S_n} \left(\prod_{i=1}^n a_{i,g_{\sigma(i)}}\right) (-1)^\sigma \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right) \\
&= \sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}} \left(\sum_{\sigma\in S_n} \left(\prod_{i=1}^n a_{i,g_{\sigma(i)}}\right) (-1)^\sigma\right) \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right). \qquad (392)
\end{aligned}$$

But every $(g_1, g_2, \ldots, g_n) \in \mathbf{I}$ satisfies $\sum_{\sigma\in S_n} \left(\prod_{i=1}^n a_{i,g_{\sigma(i)}}\right) (-1)^\sigma = \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right)$

[232]*Proof of (391):* Let $(k_1, k_2, \ldots, k_n) \in [m]^n$ and $\sigma \in S_n$. We have $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B = \left(b_{k_i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (as we have seen in one of the previous footnotes) and $\operatorname{rows}_{k_{\sigma(1)},k_{\sigma(2)},\ldots,k_{\sigma(n)}} B = \left(b_{k_{\sigma(i)},j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (for similar reasons). Hence, we can apply Lemma 6.17 **(a)** to σ, $\operatorname{rows}_{k_1,k_2,\ldots,k_n} B$, $b_{k_i,j}$ and $\operatorname{rows}_{k_{\sigma(1)},k_{\sigma(2)},\ldots,k_{\sigma(n)}} B$ instead of κ, B, $b_{i,j}$ and B_κ. As a consequence, we obtain

$$\det\left(\operatorname{rows}_{k_{\sigma(1)},k_{\sigma(2)},\ldots,k_{\sigma(n)}} B\right) = (-1)^\sigma \cdot \det\left(\operatorname{rows}_{k_1,k_2,\ldots,k_n} B\right).$$

This proves (391).

[233]. Hence, (392) becomes

$$\det(AB) = \sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}} \underbrace{\left(\sum_{\sigma\in S_n}\left(\prod_{i=1}^{n} a_{i,g_{\sigma(i)}}\right)(-1)^{\sigma}\right)}_{=\det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right)} \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right)$$
$$= \sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right) \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right). \quad (393)$$

Finally, we recall that $\mathbf{I}$ was defined as the set

$$\left\{(k_1,k_2,\ldots,k_n)\in[m]^n \mid k_1<k_2<\cdots<k_n\right\}.$$

Thus, summing over all $(g_1,g_2,\ldots,g_n)\in\mathbf{I}$ means the same as summing over all $(g_1,g_2,\ldots,g_n)\in[m]^n$ satisfying $g_1<g_2<\cdots<g_n$. In other words,

$$\sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}} = \sum_{\substack{(g_1,g_2,\ldots,g_n)\in[m]^n;\\ g_1<g_2<\cdots<g_n}} = \sum_{1\leq g_1<g_2<\cdots<g_n\leq m}$$

(an equality between summation signs – hopefully its meaning is obvious). Hence, (393) becomes

$$\det(AB) = \sum_{1\leq g_1<g_2<\cdots<g_n\leq m} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right)\cdot\det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_n} B\right).$$

This proves Theorem 6.32. □

6.6. Prelude to Laplace expansion

Next we shall show a fact which will allow us to compute some determinants by induction:

[233] *Proof.* Let $(g_1,g_2,\ldots,g_n)\in\mathbf{I}$. We have $A=(a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$. Thus, the definition of $\operatorname{cols}_{g_1,g_2,\ldots,g_n} A$ yields

$$\operatorname{cols}_{g_1,g_2,\ldots,g_n} A = \left(a_{i,g_y}\right)_{1\leq i\leq n,\ 1\leq y\leq n} = \left(a_{i,g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$$

(here, we renamed the index y as j). Hence, (341) (applied to $\operatorname{cols}_{g_1,g_2,\ldots,g_n} A$ and a_{i,g_j} instead of A and $a_{i,j}$) yields

$$\det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_n} A\right) = \sum_{\sigma\in S_n}(-1)^{\sigma}\prod_{i=1}^{n} a_{i,g_{\sigma(i)}} = \sum_{\sigma\in S_n}\left(\prod_{i=1}^{n} a_{i,g_{\sigma(i)}}\right)(-1)^{\sigma},$$

qed.

Theorem 6.43. Let n be a positive integer. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. Assume that

$$a_{n,j} = 0 \qquad \text{for every } j \in \{1, 2, \ldots, n-1\}. \tag{394}$$

Then, $\det A = a_{n,n} \cdot \det \left((a_{i,j})_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \right)$.

The assumption (394) says that the last row of the matrix A consists entirely of zeroes, apart from its last entry $a_{n,n}$ (which may and may not be 0). Theorem 6.43 states that, under this assumption, the determinant can be obtained by multiplying this last entry $a_{n,n}$ with the determinant of the $(n-1) \times (n-1)$-matrix obtained by removing both the last row and the last column from A. For example, for $n = 3$, Theorem 6.43 states that

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & g \end{pmatrix} = g \det \begin{pmatrix} a & b \\ d & e \end{pmatrix}.$$

Theorem 6.43 is a particular case of *Laplace expansion*, which is a general recursive formula for the determinants that we will encounter further below. But Theorem 6.43 already has noticeable applications of its own, which is why I have chosen to start with this particular case.

The proof of Theorem 6.43 essentially relies on the following fact:

Lemma 6.44. Let n be a positive integer. Let $(a_{i,j})_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1}$ be an $(n-1) \times (n-1)$-matrix. Then,

$$\sum_{\substack{\sigma \in S_n; \\ \sigma(n) = n}} (-1)^{\sigma} \prod_{i=1}^{n-1} a_{i,\sigma(i)} = \det \left((a_{i,j})_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \right).$$

Proof of Lemma 6.44. We define a subset T of S_n by

$$T = \{ \tau \in S_n \mid \tau(n) = n \}.$$

(In other words, T is the set of all $\tau \in S_n$ such that if we write τ in one-line notation, then τ ends with an n.)

Now, we shall construct two mutually inverse maps between S_{n-1} and T.

For every $\sigma \in S_{n-1}$, we define a map $\widehat{\sigma} : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ by setting

$$\widehat{\sigma}(i) = \begin{cases} \sigma(i), & \text{if } i < n; \\ n, & \text{if } i = n \end{cases} \qquad \text{for every } i \in \{1, 2, \ldots, n\}.$$

[234] It is straightforward to see that this map $\widehat{\sigma}$ is well-defined and belongs to T. Thus, we can define a map $\Phi : S_{n-1} \to T$ by setting

$$\Phi(\sigma) = \widehat{\sigma} \qquad \text{for every } \sigma \in S_{n-1}.$$

Loosely speaking, for every $\sigma \in S_{n-1}$, the permutation $\Phi(\sigma) = \widehat{\sigma} \in T$ is obtained by writing σ in one-line notation and appending n on its right. For example, if $n = 4$ and if $\sigma \in S_3$ is the permutation that is written as $(2,3,1)$ in one-line notation, then $\Phi(\sigma) = \widehat{\sigma}$ is the permutation that is written as $(2,3,1,4)$ in one-line notation.

On the other hand, for every $\gamma \in T$, we define a map $\overline{\gamma} : \{1,2,\ldots,n-1\} \to \{1,2,\ldots,n-1\}$ by setting

$$\overline{\gamma}(i) = \gamma(i) \qquad \text{for every } i \in \{1,2,\ldots,n-1\}.$$

It is straightforward to see that this map $\overline{\gamma}$ is well-defined and belongs to S_{n-1}. Hence, we can define a map $\Psi : T \to S_{n-1}$ by setting

$$\Psi(\gamma) = \overline{\gamma} \qquad \text{for every } \gamma \in T.$$

Loosely speaking, for every $\gamma \in T$, the permutation $\Psi(\gamma) = \overline{\gamma} \in S_{n-1}$ is obtained by writing γ in one-line notation and removing the n (which is the rightmost entry in the one-line notation, because $\gamma(n) = n$). For example, if $n = 4$ and if $\gamma \in S_4$ is the permutation that is written as $(2,3,1,4)$ in one-line notation, then $\Psi(\gamma) = \overline{\gamma}$ is the permutation that is written as $(2,3,1)$ in one-line notation.

The maps Φ and Ψ are mutually inverse.[235] Thus, the map Φ is a bijection.

It is fairly easy to see that every $\sigma \in S_{n-1}$ satisfies

$$(-1)^{\widehat{\sigma}} = (-1)^{\sigma} \tag{395}$$

[236] and

$$\prod_{i=1}^{n-1} a_{i,\widehat{\sigma}(i)} = \prod_{i=1}^{n-1} a_{i,\sigma(i)} \tag{396}$$

[234] Note that if we use Definition 5.55, then this map $\widehat{\sigma}$ is exactly the map $\sigma^{(\{1,2,\ldots,n-1\}\to\{1,2,\ldots,n\})}$.

[235] This should be clear enough from the descriptions we gave using one-line notation. A formal proof is straightforward.

[236] *Proof of (395):* Let $\sigma \in S_{n-1}$. We want to prove that $(-1)^{\widehat{\sigma}} = (-1)^{\sigma}$. It is clearly sufficient to show that $\ell(\widehat{\sigma}) = \ell(\sigma)$ (because $(-1)^{\widehat{\sigma}} = (-1)^{\ell(\widehat{\sigma})}$ and $(-1)^{\sigma} = (-1)^{\ell(\sigma)}$). In order to do so, it is sufficient to show that the inversions of $\widehat{\sigma}$ are precisely the inversions of σ (because $\ell(\widehat{\sigma})$ is the number of inversions of $\widehat{\sigma}$, whereas $\ell(\sigma)$ is the number of inversions of σ).

If (i,j) is an inversion of σ, then (i,j) is an inversion of $\widehat{\sigma}$ (because if (i,j) is an inversion of σ, then both i and j are $< n$, and thus the definition of $\widehat{\sigma}$ yields $\widehat{\sigma}(i) = \sigma(i)$ and $\widehat{\sigma}(j) = \sigma(j)$). In other words, every inversion of σ is an inversion of $\widehat{\sigma}$.

On the other hand, let (u,v) be an inversion of $\widehat{\sigma}$. We shall prove that (u,v) is an inversion of σ.

Indeed, (u,v) is an inversion of $\widehat{\sigma}$. In other words, (u,v) is a pair of integers satisfying $1 \leq u < v \leq n$ and $\widehat{\sigma}(u) > \widehat{\sigma}(v)$.

If we had $v = n$, then we would have $\widehat{\sigma}(u) > \widehat{\sigma}\left(\underbrace{v}_{=n}\right) = \widehat{\sigma}(n) = n$ (by the definition of

[237].

Now,

$$\underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(n)=n}}}_{\substack{=\sum\limits_{\sigma\in\{\tau\in S_n\ \mid\ \tau(n)=n\}}=\sum\limits_{\sigma\in T}\\ \text{(since } \{\tau\in S_n\ \mid\ \tau(n)=n\}=T)}} (-1)^{\sigma}\prod_{i=1}^{n-1}a_{i,\sigma(i)}$$

$$= \sum_{\sigma\in T}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{i,\sigma(i)} = \sum_{\sigma\in S_{n-1}}\underbrace{(-1)^{\Phi(\sigma)}\prod_{i=1}^{n-1}a_{i,(\Phi(\sigma))(i)}}_{\substack{=(-1)^{\widehat{\sigma}}\prod\limits_{i=1}^{n-1}a_{i,\widehat{\sigma}(i)}\\ \text{(since } \Phi(\sigma)=\widehat{\sigma})}}$$

$$\left(\begin{array}{c}\text{here, we have substituted } \Phi(\sigma) \text{ for } \sigma \text{ in the sum,}\\ \text{since the map } \Phi : S_{n-1}\to T \text{ is a bijection}\end{array}\right)$$

$$= \sum_{\sigma\in S_{n-1}}\underbrace{(-1)^{\widehat{\sigma}}}_{\substack{=(-1)^{\sigma}\\ \text{(by (395))}}}\underbrace{\prod_{i=1}^{n-1}a_{i,\widehat{\sigma}(i)}}_{\substack{=\prod\limits_{i=1}^{n-1}a_{i,\sigma(i)}\\ \text{(by (396))}}} = \sum_{\sigma\in S_{n-1}}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{i,\sigma(i)}.$$

Compared with

$$\det\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right) = \sum_{\sigma\in S_{n-1}}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{i,\sigma(i)}$$
$$\left(\begin{array}{c}\text{by (341), applied to } n-1 \text{ and}\\ \left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1} \text{ instead of } n \text{ and } A\end{array}\right),$$

this yields

$$\sum_{\substack{\sigma\in S_n;\\ \sigma(n)=n}}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{i,\sigma(i)} = \det\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right).$$

$\widehat{\sigma}$), which would contradict $\widehat{\sigma}(u)\in\{1,2,\ldots,n\}$. Thus, we cannot have $v=n$. We therefore have $v<n$, so that $v\leq n-1$. Now, $1\leq u<v\leq n-1$. Thus, both $\sigma(u)$ and $\sigma(v)$ are well-defined. The definition of $\widehat{\sigma}$ yields $\widehat{\sigma}(u)=\sigma(u)$ (since $u\leq n-1<n$) and $\widehat{\sigma}(v)=\sigma(v)$ (since $v\leq n-1<n$), so that $\sigma(u)=\widehat{\sigma}(u)>\widehat{\sigma}(v)=\sigma(v)$. Thus, (u,v) is a pair of integers satisfying $1\leq u<v\leq n-1$ and $\sigma(u)>\sigma(v)$. In other words, (u,v) is an inversion of σ.

We thus have shown that every inversion of $\widehat{\sigma}$ is an inversion of σ. Combining this with the fact that every inversion of σ is an inversion of $\widehat{\sigma}$, we thus conclude that the inversions of $\widehat{\sigma}$ are precisely the inversions of σ. As we have already said, this finishes the proof of (395).

[237] *Proof of (396):* Let $\sigma\in S_{n-1}$. The definition of $\widehat{\sigma}$ yields $\widehat{\sigma}(i)=\sigma(i)$ for every $i\in\{1,2,\ldots,n-1\}$. Thus, $a_{i,\widehat{\sigma}(i)}=a_{i,\sigma(i)}$ for every $i\in\{1,2,\ldots,n-1\}$. Hence, $\prod\limits_{i=1}^{n-1}\underbrace{a_{i,\widehat{\sigma}(i)}}_{=a_{i,\sigma(i)}}=\prod\limits_{i=1}^{n-1}a_{i,\sigma(i)}$, qed.

This proves Lemma 6.44. □

Proof of Theorem 6.43. Every permutation $\sigma \in S_n$ satisfying $\sigma(n) \neq n$ satisfies

$$a_{n,\sigma(n)} = 0 \tag{397}$$

[238].

From (341), we obtain

$$\begin{aligned}
\det A &= \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^{n} a_{i,\sigma(i)}}_{= \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) a_{n,\sigma(n)}} = \sum_{\sigma \in S_n} (-1)^\sigma \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) a_{n,\sigma(n)} \\
&= \sum_{\substack{\sigma \in S_n; \\ \sigma(n)=n}} (-1)^\sigma \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) \underbrace{a_{n,\sigma(n)}}_{\substack{=a_{n,n} \\ (\text{since } \sigma(n)=n)}} + \sum_{\substack{\sigma \in S_n; \\ \sigma(n)\neq n}} (-1)^\sigma \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) \underbrace{a_{n,\sigma(n)}}_{\substack{=0 \\ (\text{by (397)})}} \\
&\qquad (\text{since every } \sigma \in S_n \text{ satisfies either } \sigma(n) = n \text{ or } \sigma(n) \neq n \text{ (but not both)}) \\
&= \sum_{\substack{\sigma \in S_n; \\ \sigma(n)=n}} (-1)^\sigma \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) a_{n,n} + \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma(n)\neq n}} (-1)^\sigma \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) 0}_{=0} \\
&= \sum_{\substack{\sigma \in S_n; \\ \sigma(n)=n}} (-1)^\sigma \left(\prod_{i=1}^{n-1} a_{i,\sigma(i)}\right) a_{n,n} = a_{n,n} \cdot \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma(n)=n}} (-1)^\sigma \prod_{i=1}^{n-1} a_{i,\sigma(i)}}_{\substack{=\det\left(\left(a_{i,j}\right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1}\right) \\ (\text{by Lemma 6.44})}} \\
&= a_{n,n} \cdot \det\left(\left(a_{i,j}\right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1}\right).
\end{aligned}$$

This proves Theorem 6.43. □

Let us finally state an analogue of Theorem 6.43 in which the last column (rather than the last row) is required to consist mostly of zeroes:

Corollary 6.45. Let n be a positive integer. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. Assume that

$$a_{i,n} = 0 \qquad \text{for every } i \in \{1, 2, \ldots, n-1\}. \tag{398}$$

Then, $\det A = a_{n,n} \cdot \det\left(\left(a_{i,j}\right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1}\right)$.

[238] *Proof of (397):* Let $\sigma \in S_n$ be a permutation satisfying $\sigma(n) \neq n$. Since $\sigma(n) \in \{1, 2, \ldots, n\}$ and $\sigma(n) \neq n$, we have $\sigma(n) \in \{1, 2, \ldots, n\} \setminus \{n\} = \{1, 2, \ldots, n-1\}$. Hence, (394) (applied to $j = \sigma(n)$) shows that $a_{n,\sigma(n)} = 0$, qed.

Proof of Corollary 6.45. We have $n-1 \in \mathbb{N}$ (since n is a positive integer).

We have $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, and thus $A^T = \left(a_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of the transpose matrix A^T). Also, for every $j \in \{1,2,\ldots,n-1\}$, we have $a_{j,n} = 0$ (by (398), applied to $i = j$). Thus, Theorem 6.43 (applied to A^T and $a_{j,i}$ instead of A and $a_{i,j}$) yields

$$\det\left(A^T\right) = a_{n,n} \cdot \det\left(\left(a_{j,i}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right). \qquad (399)$$

But Exercise 6.4 shows that $\det\left(A^T\right) = \det A$. Thus, $\det A = \det\left(A^T\right)$. Also, the definition of the transpose of a matrix shows that $\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)^T = \left(a_{j,i}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}$. Thus,

$$\det\left(\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)^T\right) = \det\left(\left(a_{j,i}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right).$$

Comparing this with

$$\det\left(\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)^T\right) = \det\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)$$

(by Exercise 6.4, applied to $n-1$ and $\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}$ instead of n and A), we obtain

$$\det\left(\left(a_{j,i}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right) = \det\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right).$$

Now,

$$\begin{aligned}\det A = \det\left(A^T\right) &= a_{n,n} \cdot \underbrace{\det\left(\left(a_{j,i}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)}_{=\det\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)} \qquad \text{(by (399))} \\ &= a_{n,n} \cdot \det\left(\left(a_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right).\end{aligned}$$

This proves Corollary 6.45. □

6.7. The Vandermonde determinant

6.7.1. The statement

An example for an application of Theorem 6.43 is the famous *Vandermonde determinant*:

Theorem 6.46. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Then:

(a) We have

$$\det\left(\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

(b) We have

$$\det\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)=\prod_{1\leq i<j\leq n}\left(x_i-x_j\right).$$

(c) We have

$$\det\left(\left(x_i^{j-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)=\prod_{1\leq j<i\leq n}\left(x_i-x_j\right).$$

(d) We have

$$\det\left(\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)=\prod_{1\leq j<i\leq n}\left(x_i-x_j\right).$$

Remark 6.47. For $n=4$, the four matrices appearing in Theorem 6.46 are

$$\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=\begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1\\ x_2^3 & x_2^2 & x_2 & 1\\ x_3^3 & x_3^2 & x_3 & 1\\ x_4^3 & x_4^2 & x_4 & 1\end{pmatrix},$$

$$\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=\begin{pmatrix} x_1^3 & x_2^3 & x_3^3 & x_4^3\\ x_1^2 & x_2^2 & x_3^2 & x_4^2\\ x_1 & x_2 & x_3 & x_4\\ 1 & 1 & 1 & 1\end{pmatrix},$$

$$\left(x_i^{j-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=\begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3\\ 1 & x_2 & x_2^2 & x_2^3\\ 1 & x_3 & x_3^2 & x_3^3\\ 1 & x_4 & x_4^2 & x_4^3\end{pmatrix},$$

$$\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=\begin{pmatrix} 1 & 1 & 1 & 1\\ x_1 & x_2 & x_3 & x_4\\ x_1^2 & x_2^2 & x_3^2 & x_4^2\\ x_1^3 & x_2^3 & x_3^3 & x_4^3\end{pmatrix}.$$

It is clear that the second of these four matrices is the transpose of the first; the fourth is the transpose of the third; and the fourth is obtained from the second by rearranging the rows in opposite order. Thus, the four parts of Theorem 6.46 are rather easily seen to be equivalent. (We shall prove part **(a)** and derive the others from it.) Nevertheless it is useful to have seen them all.

Theorem 6.46 is a classical result (known as the *Vandermonde determinant*, although it is unclear whether it has been proven by Vandermonde): Almost all texts on linear algebra mention it (or, rather, at least one of its four parts), although some only prove it in lesser generality. It is a fundamental result that has various applications to abstract algebra, number theory, coding theory, combinatorics and numerical mathematics.

Theorem 6.46 has many known proofs[239]. My favorite proof (of Theorem 6.46 **(c)** only, but as I said the other parts are easily seen to be equivalent) is given in [Grinbe10, Theorem 1]. In these notes, I will show two others.

6.7.2. A proof by induction

The first proof I shall present has the advantage of demonstrating how Theorem 6.43 can be used (together with induction) in computing determinants.

Example 6.48. Let $x, y, z \in \mathbb{K}$. Let $A = \begin{pmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{pmatrix}$. Then, (343) shows that

$$\begin{aligned} \det A &= 1yz^2 + xy^2 \cdot 1 + x^2 \cdot 1z - 1y^2z - x \cdot 1z^2 - x^2y \cdot 1 \\ &= yz^2 + xy^2 + x^2z - y^2z - xz^2 - x^2y \\ &= yz(z-y) + zx(x-z) + xy(y-x). \end{aligned} \tag{400}$$

On the other hand, Theorem 6.46 **(c)** (applied to $n = 3$, $x_1 = x$, $x_2 = y$ and $x_3 = z$) yields $\det A = (y-x)(z-x)(z-y)$. Compared with (400), this yields

$$(y-x)(z-x)(z-y) = yz(z-y) + zx(x-z) + xy(y-x). \tag{401}$$

You might have encountered this curious identity as a trick of use in contest problems. When x, y, z are three distinct complex numbers, we can divide (401) by $(y-x)(z-x)(z-y)$, and obtain

$$1 = \frac{yz}{(y-x)(z-x)} + \frac{zx}{(z-y)(x-y)} + \frac{xy}{(x-z)(y-z)}.$$

Before we prove Theorem 6.46, let us see (in greater generality) what happens to the determinant of a matrix if we rearrange the rows in opposite order:

Lemma 6.49. Let $n \in \mathbb{N}$. Let $(a_{i,j})_{1 \le i \le n,\ 1 \le j \le n}$ be an $n \times n$-matrix. Then,

$$\det\left(\left(a_{n+1-i,j}\right)_{1 \le i \le n,\ 1 \le j \le n} \right) = (-1)^{n(n-1)/2} \det\left(\left(a_{i,j}\right)_{1 \le i \le n,\ 1 \le j \le n} \right).$$

[239] For four combinatorial proofs, see [Gessel79], [Aigner07, §5.3], [Loehr11, §12.9] and [BenDre07]. (Specifically, [Gessel79] and [BenDre07] prove Theorem 6.46 **(c)**, whereas [Aigner07, §5.3] and [Loehr11, §12.9] prove Theorem 6.46 **(b)**. But as we will see, the four parts of Theorem 6.46 are easily seen to be equivalent to each other.)

Proof of Lemma 6.49. Let $[n]$ denote the set $\{1,2,\ldots,n\}$. Define a permutation w_0 in S_n as in Exercise 5.11. In the solution of Exercise 5.11, we have shown that $(-1)^{w_0} = (-1)^{n(n-1)/2}$.

Now, we can apply Lemma 6.17 **(a)** to $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, w_0 and $\left(a_{w_0(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ instead of B, κ and B_κ. As a result, we obtain

$$\det\left(\left(a_{w_0(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \underbrace{(-1)^{w_0}}_{=(-1)^{n(n-1)/2}} \cdot \det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = (-1)^{n(n-1)/2} \det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right). \tag{402}$$

But $w_0(i) = n+1-i$ for every $i \in \{1,2,\ldots,n\}$ (by the definition of w_0). Thus, (402) rewrites as $\det\left(\left(a_{n+1-i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = (-1)^{n(n-1)/2} \det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$. This proves Lemma 6.49. $\square$

First proof of Theorem 6.46. **(a)** For every $u \in \{0,1,\ldots,n\}$, let A_u be the $u\times u$-matrix $\left(x_i^{u-j}\right)_{1\leq i\leq u,\ 1\leq j\leq u}$.

Now, let us show that

$$\det(A_u) = \prod_{1\leq i<j\leq u} \left(x_i - x_j\right) \tag{403}$$

for every $u \in \{0,1,\ldots,n\}$.

[*Proof of (403):* We will prove (403) by induction over u:

Induction base: The matrix A_0 is a 0×0-matrix and thus has determinant $\det(A_0) = 1$. On the other hand, the product $\prod_{1\leq i<j\leq 0} \left(x_i - x_j\right)$ is an empty product (i.e., a product of 0 elements of $\mathbb{K}$) and thus equals 1 as well. Hence, both $\det(A_0)$ and $\prod_{1\leq i<j\leq 0} \left(x_i - x_j\right)$ equal 1. Thus, $\det(A_0) = \prod_{1\leq i<j\leq 0} \left(x_i - x_j\right)$. In other words, (403) holds for $u = 0$. The induction base is thus complete.

Induction step: Let $U \in \{1,2,\ldots,n\}$. Assume that (403) holds for $u = U-1$. We need to prove that (403) holds for $u = U$.

Recall that $A_U = \left(x_i^{U-j}\right)_{1\leq i\leq U,\ 1\leq j\leq U}$ (by the definition of A_U).

For every $(i,j) \in \{1,2,\ldots,U\}^2$, define $b_{i,j} \in \mathbb{K}$ by

$$b_{i,j} = \begin{cases} x_i^{U-j} - x_U x_i^{U-j-1}, & \text{if } j < U; \\ 1, & \text{if } j = U \end{cases}.$$

Let B be the $U \times U$-matrix $\left(b_{i,j}\right)_{1\leq i\leq U,\ 1\leq j\leq U}$. For example, if $U = 4$, then

$$A_U = \begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ x_3^3 & x_3^2 & x_3 & 1 \\ x_4^3 & x_4^2 & x_4 & 1 \end{pmatrix} \qquad \text{and}$$

$$B = \begin{pmatrix} x_1^3 - x_4x_1^2 & x_1^2 - x_4x_1 & x_1 - x_4 & 1 \\ x_2^3 - x_4x_2^2 & x_2^2 - x_4x_2 & x_2 - x_4 & 1 \\ x_3^3 - x_4x_3^2 & x_3^2 - x_4x_3 & x_3 - x_4 & 1 \\ x_4^3 - x_4x_4^2 & x_4^2 - x_4x_4 & x_4 - x_4 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} x_1^3 - x_4x_1^2 & x_1^2 - x_4x_1 & x_1 - x_4 & 1 \\ x_2^3 - x_4x_2^2 & x_2^2 - x_4x_2 & x_2 - x_4 & 1 \\ x_3^3 - x_4x_3^2 & x_3^2 - x_4x_3 & x_3 - x_4 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We claim that $\det B = \det(A_U)$. Indeed, here are two ways to prove this:

First proof of $\det B = \det(A_U)$*:* Exercise 6.8 **(b)** shows that the determinant of a $U \times U$-matrix does not change if we subtract a multiple of one of its columns from another column. Now, let us subtract x_U times the 2-nd column of A_U from the 1-st column, then subtract x_U times the 3-rd column of the resulting matrix from the 2-nd column, and so on, all the way until we finally subtract x_U times the U-th column of the matrix from the $(U-1)$-st column[240]. The resulting matrix is B (according to our definition of B). Thus, $\det B = \det(A_U)$ (since our subtractions never change the determinant). This proves $\det B = \det(A_U)$.

Second proof of $\det B = \det(A_U)$*:* Here is another way to prove that $\det B = \det(A_U)$, with some less handwaving.

For every $(i,j) \in \{1,2,\ldots,U\}^2$, we define $c_{i,j} \in \mathbb{K}$ by

$$c_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ -x_U, & \text{if } i = j+1; \\ 0, & \text{otherwise} \end{cases}.$$

Let C be the $U \times U$-matrix $\left(c_{i,j}\right)_{1\leq i\leq U,\ 1\leq j\leq U}$.

For example, if $U = 4$, then

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -x_4 & 1 & 0 & 0 \\ 0 & -x_4 & 1 & 0 \\ 0 & 0 & -x_4 & 1 \end{pmatrix}.$$

[240]So, all in all, we subtract the x_U-multiple of each column from its neighbor to its left, but the order in which we are doing it (namely, from left to right) is important: It means that the column we are subtracting is unchanged from A_U. (If we would be doing these subtractions from right to left instead, then the columns to be subtracting would be changed by the preceding steps.)

The matrix C is lower-triangular, and thus Exercise 6.3 shows that its determinant is $\det C = \underbrace{c_{1,1}}_{=1} \underbrace{c_{2,2}}_{=1} \cdots \underbrace{c_{U,U}}_{=1} = 1$.

On the other hand, it is easy to see that $B = A_U C$ (check this!). Thus, Theorem 6.23 yields $\det B = \det(A_U) \cdot \underbrace{\det C}_{=1} = \det(A_U)$. So we have proven $\det B = \det(A_U)$ again.

[*Remark:* It is instructive to compare the two proofs of $\det B = \det(A_U)$ given above. They are close kin, although they might look different at first. In the first proof, we argued that B can be obtained from A_U by subtracting multiples of some columns from others; in the second, we argued that $B = A_U C$ for a specific lower-triangular matrix C. But a look at the matrix C makes it clear that multiplying a $U \times U$-matrix with C on the right (i.e., transforming a $U \times U$-matrix X into the matrix XC) is tantamount to subtracting multiples of some columns from others, in the way we did it to A_U to obtain B. So the main difference between the two proofs is that the first proof used a step-by-step procedure to obtain B from A_U, whereas the second proof obtained B from A_U by a single-step operation (namely, multiplication by a matrix C).]

Next, we observe that for every $j \in \{1, 2, \ldots, U-1\}$, we have

$$\begin{aligned} b_{U,j} &= \begin{cases} x_U^{U-j} - x_U x_U^{U-j-1}, & \text{if } j < U; \\ 1, & \text{if } j = U \end{cases} \qquad (\text{by the definition of } b_{U,j}) \\ &= x_U^{U-j} - \underbrace{x_U x_U^{U-j-1}}_{= x_U^{(U-j-1)+1} = x_U^{U-j}} \qquad (\text{since } j < U \ (\text{since } j \in \{1, 2, \ldots, U-1\})) \\ &= x_U^{U-j} - x_U^{U-j} = 0. \end{aligned}$$

Hence, Theorem 6.43 (applied to U, B and $b_{i,j}$ instead of n, A and $a_{i,j}$) yields

$$\det B = b_{U,U} \cdot \det\left(\left(b_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1} \right). \tag{404}$$

Let B' denote the $(U-1) \times (U-1)$-matrix $\left(b_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}$.

The definition of $b_{U,U}$ yields

$$\begin{aligned} b_{U,U} &= \begin{cases} x_U^{U-U} - x_U x_U^{U-U-1}, & \text{if } U < U; \\ 1, & \text{if } U = U \end{cases} \qquad (\text{by the definition of } b_{U,U}) \\ &= 1 \qquad (\text{since } U = U). \end{aligned}$$

Thus, (404) becomes

$$\det B = \underbrace{b_{U,U}}_{=1} \cdot \det\left(\underbrace{\left(b_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}}_{=B'} \right) = \det\left(B' \right).$$

Compared with $\det B = \det(A_U)$, this yields

$$\det(A_U) = \det(B'). \tag{405}$$

Now, let us take a closer look at B'. Indeed, every $(i,j) \in \{1,2,\ldots,U-1\}^2$ satisfies

$$\begin{aligned} b_{i,j} &= \begin{cases} x_i^{U-j} - x_U x_i^{U-j-1}, & \text{if } j < U; \\ 1, & \text{if } j = U \end{cases} \qquad \left(\text{by the definition of } b_{i,j}\right) \\ &= \underbrace{x_i^{U-j}}_{=x_i x_i^{U-j-1}} - x_U x_i^{U-j-1} \qquad \left(\begin{array}{c} \text{since } j < U \text{ (since } j \in \{1,2,\ldots,U-1\} \\ \text{(since } (i,j) \in \{1,2,\ldots,U-1\}^2)) \end{array}\right) \\ &= x_i x_i^{U-j-1} - x_U x_i^{U-j-1} = (x_i - x_U) \underbrace{x_i^{U-j-1}}_{=x_i^{(U-1)-j}} = (x_i - x_U) x_i^{(U-1)-j}. \end{aligned} \tag{406}$$

Hence,

$$B' = \left(\underbrace{b_{i,j}}_{\substack{=(x_i - x_U) x_i^{(U-1)-j} \\ \text{(by (406))}}} \right)_{1\le i\le U-1,\ 1\le j\le U-1} = \left((x_i - x_U) x_i^{(U-1)-j} \right)_{1\le i\le U-1,\ 1\le j\le U-1}. \tag{407}$$

On the other hand, the definition of A_{U-1} yields

$$A_{U-1} = \left(x_i^{(U-1)-j} \right)_{1\le i\le U-1,\ 1\le j\le U-1}. \tag{408}$$

Now, we claim that

$$\det(B') = \det(A_{U-1}) \cdot \prod_{i=1}^{U-1} (x_i - x_U). \tag{409}$$

Indeed, here are two ways to prove this:

First proof of (409): Comparing the formulas (407) and (408), we see that the matrix B' is obtained from the matrix A_{U-1} by multiplying the first row by $x_1 - x_U$, the second row by $x_2 - x_U$, and so on, and finally the $(U-1)$-st row by $x_{U-1} - x_U$. But every time we multiply a row of a $(U-1) \times (U-1)$-matrix by some scalar $\lambda \in \mathbb{K}$, the determinant of the matrix gets multiplied by λ (because of Exercise 6.7 **(g)**). Hence, the determinant of B' is obtained from that of A_{U-1} by first multiplying by $x_1 - x_U$, then multiplying by $x_2 - x_U$, and so on, and finally multiplying with $x_{U-1} - x_U$. In other words,

$$\det(B') = \det(A_{U-1}) \cdot \prod_{i=1}^{U-1} (x_i - x_U).$$

This proves (409).

Second proof of (409): For every $(i,j) \in \{1,2,\ldots,U-1\}^2$, we define $d_{i,j} \in \mathbb{K}$ by

$$d_{i,j} = \begin{cases} x_i - x_U, & \text{if } i = j; \\ 0, & \text{otherwise} \end{cases}.$$

Let D be the $(U-1) \times (U-1)$-matrix $\left(d_{i,j}\right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}$.

For example, if $U = 4$, then

$$D = \left(\begin{array}{ccc} x_1 - x_4 & 0 & 0 \\ 0 & x_2 - x_4 & 0 \\ 0 & 0 & x_3 - x_4 \end{array}\right).$$

The matrix D is lower-triangular (actually, diagonal[241]), and thus Exercise 6.3 shows that its determinant is

$$\det D = (x_1 - x_U)(x_2 - x_U) \cdots (x_{U-1} - x_U) = \prod_{i=1}^{U-1} (x_i - x_U).$$

On the other hand, it is easy to see that $B' = DA_{U-1}$ (check this!). Thus, Theorem 6.23 yields

$$\det\left(B'\right) = \det D \cdot \det\left(A_{U-1}\right) = \det\left(A_{U-1}\right) \cdot \underbrace{\det D}_{= \prod_{i=1}^{U-1} (x_i - x_U)} = \det\left(A_{U-1}\right) \cdot \prod_{i=1}^{U-1} (x_i - x_U).$$

Thus, (409) is proven again.

[*Remark:* Again, our two proofs of (409) are closely related: the first one reveals B' as the result of a step-by-step process applied to A_{U-1}, while the second shows how B' can be obtained from A_{U-1} by a single multiplication. However, here (in contrast to the proofs of $\det B = \det\left(A_U\right)$), the step-by-step process involves transforming rows (not columns), and the multiplication is a multiplication from the left (we have $B' = DA_{U-1}$, not $B' = A_{U-1}D$).]

Now, (405) becomes

$$\det\left(A_U\right) = \det\left(B'\right) = \det\left(A_{U-1}\right) \cdot \prod_{i=1}^{U-1} (x_i - x_U). \tag{410}$$

But we have assumed that (403) holds for $u = U - 1$. In other words,

$$\det\left(A_{U-1}\right) = \underbrace{\prod_{1 \leq i < j \leq U-1}}_{= \prod_{j=1}^{U-1} \prod_{i=1}^{j-1}} \left(x_i - x_j\right) = \prod_{j=1}^{U-1} \prod_{i=1}^{j-1} \left(x_i - x_j\right).$$

[241] A square matrix $E = \left(e_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ is said to be *diagonal* if every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i \neq j$ satisfies $e_{i,j} = 0$. In other words, a square matrix is said to be *diagonal* if it is both upper-triangular and lower-triangular.

Hence, (410) yields

$$\det(A_U) = \underbrace{\det(A_{U-1})}_{=\prod_{j=1}^{U-1}\prod_{i=1}^{j-1}(x_i-x_j)} \cdot \prod_{i=1}^{U-1}(x_i - x_U)$$
$$= \left(\prod_{j=1}^{U-1}\prod_{i=1}^{j-1}(x_i - x_j)\right) \cdot \prod_{i=1}^{U-1}(x_i - x_U).$$

Compared with

$$\underbrace{\prod_{1\leq i<j\leq U}}_{=\prod_{j=1}^{U}\prod_{i=1}^{j-1}}(x_i - x_j) = \prod_{j=1}^{U}\prod_{i=1}^{j-1}(x_i - x_j) = \left(\prod_{j=1}^{U-1}\prod_{i=1}^{j-1}(x_i - x_j)\right) \cdot \prod_{i=1}^{U-1}(x_i - x_U)$$

(here, we have split off the factor for $j = U$ from the product),

this yields $\det(A_U) = \prod_{1\leq i<j\leq U}(x_i - x_j)$. In other words, (403) holds for $u = U$. This completes the induction step.

Now, (403) is proven by induction.]

Hence, we can apply (403) to $u = n$. As the result, we obtain $\det(A_n) = \prod_{1\leq i<j\leq n}(x_i - x_j)$. Since $A_n = \left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of A_n), this rewrites as $\det\left(\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n}(x_i - x_j)$. This proves Theorem 6.46 **(a)**.

(b) The definition of the transpose of a matrix yields $\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)^T = \left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence,

$$\det\left(\underbrace{\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)^T}_{=\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}}\right) = \det\left(\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n}(x_i - x_j)$$

(by Theorem 6.46 **(a)**). Compared with

$$\det\left(\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)^T\right) = \det\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$

(by Exercise 6.4, applied to $A = \left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$), this yields

$$\det\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

This proves Theorem 6.46 **(b)**.

(d) Applying Lemma 6.49 to $a_{i,j} = x_j^{n-i}$, we obtain

$$\det\left(\left(x_j^{n-(n+1-i)}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = (-1)^{n(n-1)/2} \underbrace{\det\left(\left(x_j^{n-i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{\substack{=\prod_{1\leq i<j\leq n}\left(x_i-x_j\right) \\ \text{(by Theorem 6.46 (b))}}}$$
$$= (-1)^{n(n-1)/2} \prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

This rewrites as

$$\det\left(\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = (-1)^{n(n-1)/2} \prod_{1\leq i<j\leq n}\left(x_i - x_j\right) \qquad (411)$$

(since every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfies $x_j^{n-(n+1-i)} = x_j^{i-1}$).

Now, in the solution to Exercise 5.11, we have shown that the number of all pairs (i,j) of integers satisfying $1 \leq i < j \leq n$ is $n(n-1)/2$. In other words,

$$\left(\text{the number of all } (i,j) \in \{1,2,\ldots,n\}^2 \text{ such that } i<j\right) = n(n-1)/2. \qquad (412)$$

Now,

$$\prod_{1\leq j<i\leq n}\left(x_i - x_j\right) = \prod_{1\leq i<j\leq n} \underbrace{\left(x_j - x_i\right)}_{=(-1)\left(x_i-x_j\right)} \qquad \left(\begin{array}{c}\text{here, we renamed the index } (j,i) \\ \text{as } (i,j) \text{ in the product}\end{array}\right)$$
$$= \prod_{1\leq i<j\leq n}\left((-1)\left(x_i - x_j\right)\right)$$
$$= \underbrace{(-1)^{\left(\text{the number of all } (i,j)\in\{1,2,\ldots,n\}^2 \text{ such that } i<j\right)}}_{\substack{=(-1)^{n(n-1)/2} \\ \text{(by (412))}}} \prod_{1\leq i<j\leq n}\left(x_i - x_j\right)$$
$$= (-1)^{n(n-1)/2} \prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

Compared with (411), this yields $\det\left(\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq j<i\leq n}\left(x_i - x_j\right)$. This proves Theorem 6.46 **(d)**.

(c) We can derive Theorem 6.46 **(c)** from Theorem 6.46 **(d)** in the same way as we derived part **(b)** from **(a)**. □

6.7.3. A proof by factoring the matrix

Next, I shall outline another proof of Theorem 6.46, which proceeds by writing the matrix $\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ as a product of a lower-triangular matrix with an upper-triangular matrix. The idea of this proof appears in [OruPhi00, Theorem 2.1], [GohKol96, Theorem 2] and [OlvSha18, proof of Lemma 5.16] (although the first two of these three sources use slightly different arguments, and the third gives no proof).

We will need several lemmas for the proof. The proofs of these lemmas are relegated to the solution of Exercise 6.15.

We begin with a definition that will be used throughout Subsection 6.7.3:

Definition 6.50. Let $k \in \mathbb{Z}$ and $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Then, we define an element $h_k(x_1, x_2, \ldots, x_n)$ by

$$h_k(x_1, x_2, \ldots, x_n) = \sum_{\substack{(a_1,a_2,\ldots,a_n)\in\mathbb{N}^n;\\ a_1+a_2+\cdots+a_n=k}} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}.$$

(Note that the sum on the right hand side of this equality is finite, because only finitely many $(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n$ satisfy $a_1 + a_2 + \cdots + a_n = k$.)

The element $h_k(x_1, x_2, \ldots, x_n)$ defined in Definition 6.50 is often called the *k-th complete homogeneous function of the n elements $x_1, x_2, \ldots, x_n$* (although, more often, this notion is reserved for a different, more abstract object of whom $h_k(x_1, x_2, \ldots, x_n)$ is just an evaluation). Let us see some examples:

Example 6.51. **(a)** If $n \in \mathbb{N}$, and if $x_1, x_2, \ldots, x_n$ are n elements of $\mathbb{K}$, then

$$h_1(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$$

and

$$h_2(x_1, x_2, \ldots, x_n) = \left(x_1^2 + x_2^2 + \cdots + x_n^2\right) + \sum_{1\leq i<j\leq n} x_i x_j.$$

For example, for $n = 3$, we obtain $h_2(x_1, x_2, x_3) = \left(x_1^2 + x_2^2 + x_3^2\right) + (x_1x_2 + x_1x_3 + x_2x_3)$.

(b) If x and y are two elements of $\mathbb{K}$, then

$$h_k(x, y) = \sum_{\substack{(a_1,a_2)\in\mathbb{N}^2;\\ a_1+a_2=k}} x^{a_1} y^{a_2} = x^k y^0 + x^{k-1} y^1 + \cdots + x^0 y^k$$

for every $k \in \mathbb{N}$.

(c) If $x \in \mathbb{K}$, then $h_k(x) = x^k$ for every $k \in \mathbb{N}$.

Lemma 6.52. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$.

(a) We have $h_k(x_1, x_2, \ldots, x_n) = 0$ for every negative integer k.

(b) We have $h_0(x_1, x_2, \ldots, x_n) = 1$.

As we have said, all lemmas in Subsection 6.7.3 will be proven in the solution to Exercise 6.15.

Three further lemmas on the $h_k(x_1, x_2, \ldots, x_n)$ will be of use:

Lemma 6.53. Let k be a positive integer. Let $x_1, x_2, \ldots, x_k$ be k elements of $\mathbb{K}$. Let $q \in \mathbb{Z}$. Then,

$$h_q(x_1, x_2, \ldots, x_k) = \sum_{r=0}^{q} x_k^r h_{q-r}(x_1, x_2, \ldots, x_{k-1}).$$

Lemma 6.54. Let k be a positive integer. Let $x_1, x_2, \ldots, x_k$ be k elements of $\mathbb{K}$. Let $q \in \mathbb{Z}$. Then,

$$h_q(x_1, x_2, \ldots, x_k) = h_q(x_1, x_2, \ldots, x_{k-1}) + x_k h_{q-1}(x_1, x_2, \ldots, x_k).$$

Lemma 6.55. Let i be a positive integer. Let $x_1, x_2, \ldots, x_i$ be i elements of $\mathbb{K}$. Let $u \in \mathbb{K}$. Then,

$$\sum_{k=1}^{i} h_{i-k}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) = u^{i-1}.$$

Next, let us introduce two matrices:

Lemma 6.56. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Define an $n \times n$-matrix $U \in \mathbb{K}^{n \times n}$ by

$$U = \left(\prod_{p=1}^{i-1} (x_j - x_p) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Then, $\det U = \prod_{1 \leq j < i \leq n} (x_i - x_j)$.

Example 6.57. If $n = 4$, then the matrix U defined in Lemma 6.56 looks as follows:

$$U = \left(\prod_{p=1}^{i-1} (x_j - x_p) \right)_{1 \le i \le 4,\ 1 \le j \le 4}$$
$$= \begin{pmatrix} \prod_{p=1}^{0} (x_1 - x_p) & \prod_{p=1}^{0} (x_2 - x_p) & \prod_{p=1}^{0} (x_3 - x_p) & \prod_{p=1}^{0} (x_4 - x_p) \\ \prod_{p=1}^{1} (x_1 - x_p) & \prod_{p=1}^{1} (x_2 - x_p) & \prod_{p=1}^{1} (x_3 - x_p) & \prod_{p=1}^{1} (x_4 - x_p) \\ \prod_{p=1}^{2} (x_1 - x_p) & \prod_{p=1}^{2} (x_2 - x_p) & \prod_{p=1}^{2} (x_3 - x_p) & \prod_{p=1}^{2} (x_4 - x_p) \\ \prod_{p=1}^{3} (x_1 - x_p) & \prod_{p=1}^{3} (x_2 - x_p) & \prod_{p=1}^{3} (x_3 - x_p) & \prod_{p=1}^{3} (x_4 - x_p) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ 0 & 0 & (x_3 - x_1)(x_3 - x_2) & (x_4 - x_1)(x_4 - x_2) \\ 0 & 0 & 0 & (x_4 - x_1)(x_4 - x_2)(x_4 - x_3) \end{pmatrix}.$$

(Here, we have used the fact that if $i > j$, then the product $\prod_{p=1}^{i-1} (x_j - x_p)$ contains the factor $x_j - x_j = 0$ and thus equals 0.)

Lemma 6.58. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Define an $n \times n$-matrix $L \in \mathbb{K}^{n \times n}$ by

$$L = \left(h_{i-j} (x_1, x_2, \ldots, x_j) \right)_{1 \le i \le n,\ 1 \le j \le n}.$$

Then, $\det L = 1$.

Example 6.59. If $n = 4$, then the matrix L defined in Lemma 6.58 looks as follows:

$$L = \left(h_{i-j} (x_1, x_2, \ldots, x_j) \right)_{1 \le i \le n,\ 1 \le j \le n}$$
$$= \begin{pmatrix} h_0 (x_1) & h_{-1} (x_1, x_2) & h_{-2} (x_1, x_2, x_3) & h_{-3} (x_1, x_2, x_3, x_4) \\ h_1 (x_1) & h_0 (x_1, x_2) & h_{-1} (x_1, x_2, x_3) & h_{-2} (x_1, x_2, x_3, x_4) \\ h_2 (x_2) & h_1 (x_1, x_2) & h_0 (x_1, x_2, x_3) & h_{-1} (x_1, x_2, x_3, x_4) \\ h_3 (x_3) & h_2 (x_1, x_2) & h_1 (x_1, x_2, x_3) & h_0 (x_1, x_2, x_3, x_4) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ x_1 & 1 & 0 & 0 \\ x_1^2 & x_1 + x_2 & 1 & 0 \\ x_1^3 & x_1^2 + x_1 x_2 + x_2^2 & x_1 + x_2 + x_3 & 1 \end{pmatrix}.$$

(The fact that the diagonal entries are 1 is a consequence of Lemma 6.52 **(b)**, and the fact that the entries above the diagonal are 0 is a consequence of Lemma 6.52 **(a)**.)

Lemma 6.60. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let L be the $n \times n$-matrix defined in Lemma 6.58. Let U be the $n \times n$-matrix defined in Lemma 6.56. Then,

$$\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = LU.$$

Exercise 6.15. Prove Lemma 6.52, Lemma 6.53, Lemma 6.54, Lemma 6.55, Lemma 6.56, Lemma 6.58 and Lemma 6.60.

Now, we can prove Theorem 6.46 again:

Second proof of Theorem 6.46. **(d)** Let L be the $n \times n$-matrix defined in Lemma 6.58. Let U be the $n \times n$-matrix defined in Lemma 6.56. Then, Lemma 6.60 yields $\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = LU$. Hence,

$$\det\left(\underbrace{\left(x_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}}_{=LU}\right) = \det(LU) = \underbrace{\det L}_{\substack{=1 \\ \text{(by Lemma 6.58)}}} \cdot \underbrace{\det U}_{\substack{=\prod_{1\leq j<i\leq n}(x_i-x_j) \\ \text{(by Lemma 6.56)}}}$$

$$\left(\begin{array}{c}\text{by Theorem 6.23, applied to } L \text{ and } U \\ \text{instead of } A \text{ and } B\end{array}\right)$$

$$= 1 \cdot \prod_{1\leq j<i\leq n}\left(x_i - x_j\right) = \prod_{1\leq j<i\leq n}\left(x_i - x_j\right).$$

This proves Theorem 6.46 **(d)**.

Now, it remains to prove parts **(a)**, **(b)** and **(c)** of Theorem 6.46. This is fairly easy: Back in our First proof of Theorem 6.46, we have derived parts **(b)**, **(d)** and **(c)** from part **(a)**. By essentially the same arguments (sometimes done in reverse), we can derive parts **(a)**, **(b)** and **(c)** from part **(d)**. (We need to use the fact that $\left((-1)^{n(n-1)/2}\right)^2 = 1$.) □

6.7.4. Remarks and variations

Remark 6.61. One consequence of Theorem 6.46 is a new solution to Exercise 5.13 **(a)**:

Namely, let $n \in \mathbb{N}$ and $\sigma \in S_n$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{C}$ (or of any commutative ring). Then, Theorem 6.46 **(a)** yields

$$\det\left(\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

On the other hand, Theorem 6.46 **(a)** (applied to $x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}$ instead of $x_1, x_2, \ldots, x_n$) yields

$$\det\left(\left(x_{\sigma(i)}^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n}\left(x_{\sigma(i)} - x_{\sigma(j)}\right). \qquad (413)$$

But Lemma 6.17 **(a)** (applied to $B = \left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, $\kappa = \sigma$ and $B_\kappa = \left(x_{\sigma(i)}^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$) yields

$$\det\left(\left(x_{\sigma(i)}^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = (-1)^{\sigma}\cdot\underbrace{\det\left(\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{=\prod_{1\leq i<j\leq n}(x_i-x_j)}$$
$$= (-1)^{\sigma}\cdot\prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

Compared with (413), this yields

$$\prod_{1\leq i<j\leq n}\left(x_{\sigma(i)} - x_{\sigma(j)}\right) = (-1)^{\sigma}\cdot\prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

Thus, Exercise 5.13 **(a)** is solved. However, Exercise 5.13 **(b)** cannot be solved this way.

Exercise 6.16. Let n be a positive integer. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Prove that

$$\det\left(\left(\begin{cases} x_i^{n-j}, & \text{if } j>1; \\ x_i^n, & \text{if } j=1 \end{cases}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = (x_1 + x_2 + \cdots + x_n)\prod_{1\leq i<j\leq n}\left(x_i - x_j\right).$$

(For example, when $n = 4$, this states that

$$\det\begin{pmatrix} x_1^4 & x_1^2 & x_1 & 1 \\ x_2^4 & x_2^2 & x_2 & 1 \\ x_3^4 & x_3^2 & x_3 & 1 \\ x_4^4 & x_4^2 & x_4 & 1 \end{pmatrix}$$
$$= (x_1+x_2+x_3+x_4)(x_1-x_2)(x_1-x_3)(x_1-x_4)(x_2-x_3)(x_2-x_4)(x_3-x_4).$$

)

Remark 6.62. We can try to generalize Vandermonde's determinant. Namely, let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $a_1, a_2, \ldots, a_n$ be n nonnegative integers. Let A be the $n \times n$-matrix

$$\left(x_i^{a_j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \begin{pmatrix} x_1^{a_1} & x_1^{a_2} & \cdots & x_1^{a_n} \\ x_2^{a_1} & x_2^{a_2} & \cdots & x_2^{a_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{a_1} & x_n^{a_2} & \cdots & x_n^{a_n} \end{pmatrix}.$$

What can we say about $\det A$?

Theorem 6.46 says that if $(a_1, a_2, \ldots, a_n) = (n-1, n-2, \ldots, 0)$, then $\det A = \prod_{1 \leq i < j \leq n} (x_i - x_j)$.

Exercise 6.16 says that if $n > 0$ and $(a_1, a_2, \ldots, a_n) = (n, n-2, n-3, \ldots, 0)$, then $\det A = (x_1 + x_2 + \cdots + x_n) \prod_{1 \leq i < j \leq n} (x_i - x_j)$.

This suggests a general pattern: We would suspect that for every $(a_1, a_2, \ldots, a_n)$, there is a polynomial $P_{(a_1, a_2, \ldots, a_n)}$ in n indeterminates $X_1, X_2, \ldots, X_n$ such that

$$\det A = P_{(a_1, a_2, \ldots, a_n)}(x_1, x_2, \ldots, x_n) \cdot \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

It turns out that this is true. Moreover, this polynomial $P_{(a_1, a_2, \ldots, a_n)}$ is:

- zero if two of $a_1, a_2, \ldots, a_n$ are equal;
- homogeneous of degree $a_1 + a_2 + \cdots + a_n - \binom{n}{2}$;
- symmetric in $X_1, X_2, \ldots, X_n$.

For example,

$$\begin{aligned} P_{(n-1,n-2,\ldots,0)} &= 1; \\ P_{(n,n-2,n-3,\ldots,0)} &= \sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n; \\ P_{(n,n-1,\ldots,n-k+1,n-k-1,n-k-2,\ldots,0)} &= \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} X_{i_1} X_{i_2} \cdots X_{i_k} \\ &\qquad \text{for every } k \in \{0, 1, \ldots, n\}; \\ P_{(n+1,n-2,n-3,\ldots,0)} &= \sum_{1 \leq i \leq j \leq n} X_i X_j; \\ P_{(n+1,n-1,n-3,n-4,\ldots,0)} &= \sum_{1 \leq i < j \leq n} \left(X_i^2 X_j + X_i X_j^2\right) + 2 \sum_{1 \leq i < j < k \leq n} X_i X_j X_k. \end{aligned}$$

But this polynomial $P_{(a_1,a_2,\ldots,a_n)}$ can actually be described rather explicitly for general $(a_1, a_2, \ldots, a_n)$; it is a so-called *Schur polynomial* (at least when $a_1 > a_2 > \cdots > a_n$; otherwise it is either zero or $\pm$ a Schur polynomial). See [Stembr02, The Bi-Alternant Formula], [Stanle01, Theorem 7.15.1] or [Leeuwe06] for the details. (Notice that [Leeuwe06] uses the notation $\varepsilon(\sigma)$ for the sign of a permutation σ.) The theory of Schur polynomials shows, in particular, that all coefficients of the polynomial $P_{(a_1,a_2,\ldots,a_n)}$ have equal sign (which is positive if $a_1 > a_2 > \cdots > a_n$).

Remark 6.63. There are plenty other variations on the Vandermonde determinant. For instance, one can try replacing the powers x_i^{j-1} by binomial coefficients $\dbinom{x_i}{j-1}$ in Theorem 6.46 **(c)**, at least when these binomial coefficients are well-defined (e.g., when the $x_1, x_2, \ldots, x_n$ are complex numbers). The result is rather nice: If $x_1, x_2, \ldots, x_n$ are any n complex numbers, then

$$\prod_{1\leq i<j\leq n} \frac{x_i - x_j}{i-j} = \det\left(\left(\dbinom{x_i}{j-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

(This is proven, e.g., in [Grinbe10, Corollary 11] and [AndDos10, §9, Example 5].) This has the surprising consequence that, whenever $x_1, x_2, \ldots, x_n$ are n integers, the product $\prod\limits_{1\leq i<j\leq n} \dfrac{x_i - x_j}{i-j}$ is itself an integer (because it is the determinant of a matrix whose entries are integers). This is a nontrivial result! (A more elementary proof appears in [AndDos10, §3, Example 8].)

Another "secret integer" (i.e., rational number which turns out to be an integer for non-obvious reasons) is

$$\frac{H(a)\,H(b)\,H(c)\,H(a+b+c)}{H(b+c)\,H(c+a)\,H(a+b)}, \tag{414}$$

where a, b, c are three nonnegative integers, and where $H(n)$ (for $n \in \mathbb{N}$) denotes the *hyperfactorial* of n, defined by

$$H(n) = \prod_{k=0}^{n-1} k! = 0! \cdot 1! \cdot \cdots \cdot (n-1)!.$$

I am aware of two proofs of the fact that (414) gives an integer for every $a, b, c \in \mathbb{N}$: One proof is combinatorial, and argues that (414) is the number of *plane partitions inside an $a \times b \times c$-box* (see [Stanle01, last equality in §7.21] for a proof), or, equivalently, the number of *rhombus tilings of a hexagon with side-lengths a, b, c, a, b, c* (see [Eisenk99] for a precise statement). Another proof (see [Grinbe10, Theorem 0]) exhibits (414) as the determinant of a matrix, again using the Vandermonde determinant!

(None of the references to [Grinbe10] makes any claim of precedence; actually, I am rather sure of the opposite, i.e., that none of my proofs in [Grinbe10] are new.)

For some more exercises related to Vandermonde determinants, see [Prasol94, Chapter 1, problems 1.12–1.22]. Here comes one of them:

Exercise 6.17. Let n be a positive integer. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $y_1, y_2, \ldots, y_n$ be n elements of $\mathbb{K}$.

(a) For every $m \in \{0, 1, \ldots, n-2\}$, prove that

$$\det\left(\left((x_i + y_j)^m\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = 0.$$

(b) Prove that

$$\det\left(\left((x_i + y_j)^{n-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}(x_i - x_j)\right)\left(\prod_{1\leq i<j\leq n}(y_j - y_i)\right).$$

[Hint: Use the binomial theorem.]

(c) Let $(p_0, p_1, \ldots, p_{n-1}) \in \mathbb{K}^n$ be an n-tuple of elements of $\mathbb{K}$. Let $P(X) \in \mathbb{K}[X]$ be the polynomial $\sum\limits_{k=0}^{n-1} p_k X^k$. Prove that

$$\det\left(\left(P(x_i + y_j)\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = p_{n-1}^n \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}(x_i - x_j)\right)\left(\prod_{1\leq i<j\leq n}(y_j - y_i)\right).$$

Notice how Exercise 6.17 **(a)** generalizes Example 6.7 (for $n \geq 3$).

6.8. Invertible elements in commutative rings, and fields

We shall now interrupt our study of determinants for a moment. Let us define the notion of inverses in $\mathbb{K}$. (Recall that $\mathbb{K}$ is a commutative ring.)

Definition 6.64. Let $a \in \mathbb{K}$. Then, an element $b \in \mathbb{K}$ is said to be an *inverse* of a if it satisfies $ab = 1$ and $ba = 1$.

Of course, the two conditions $ab = 1$ and $ba = 1$ in Definition 6.64 are equivalent, since $ab = ba$ for every $a \in \mathbb{K}$ and $b \in \mathbb{K}$. Nevertheless, we have given both conditions, because this way the similarity between the inverse of an element of $\mathbb{K}$ and the inverse of a map becomes particularly clear.

For example, the element 1 of $\mathbb{Z}$ is its own inverse (since $1 \cdot 1 = 1$), and the element -1 of $\mathbb{Z}$ is its own inverse as well (since $(-1) \cdot (-1) = 1$). These elements

1 and -1 are the only elements of $\mathbb{Z}$ which have an inverse in $\mathbb{Z}$. However, in the larger commutative ring $\mathbb{Q}$, every nonzero element a has an inverse (namely, $\dfrac{1}{a}$).

Proposition 6.65. Let $a \in \mathbb{K}$. Then, there exists at most one inverse of a in $\mathbb{K}$.

Proof of Proposition 6.65. Let b and b' be any two inverses of a in $\mathbb{K}$. Since b is an inverse of a in $\mathbb{K}$, we have $ab = 1$ and $ba = 1$ (by the definition of an "inverse of a"). Since b' is an inverse of a in $\mathbb{K}$, we have $ab' = 1$ and $b'a = 1$ (by the definition of an "inverse of a"). Now, comparing $b \underbrace{ab'}_{=1} = b$ with $\underbrace{ba}_{=1} b' = b'$, we obtain $b = b'$.

Let us now forget that we fixed b and b'. We thus have shown that if b and b' are two inverses of a in $\mathbb{K}$, then $b = b'$. In other words, any two inverses of a in $\mathbb{K}$ are equal. In other words, there exists at most one inverse of a in $\mathbb{K}$. This proves Proposition 6.65. □

Definition 6.66. **(a)** An element $a \in \mathbb{K}$ is said to be *invertible* (or, more precisely, *invertible in* $\mathbb{K}$) if and only if there exists an inverse of a in $\mathbb{K}$. In this case, this inverse of a is unique (by Proposition 6.65), and thus will be called *the inverse of* a and denoted by a^{-1}.

(b) It is clear that the unity 1 of $\mathbb{K}$ is invertible (having inverse 1). Also, the product of any two invertible elements a and b of $\mathbb{K}$ is again invertible (having inverse $(ab)^{-1} = a^{-1}b^{-1}$).

(c) If a and b are two elements of $\mathbb{K}$ such that a is invertible (in $\mathbb{K}$), then we write $\dfrac{b}{a}$ (or b/a) for the product ba^{-1}. These fractions behave just as fractions of integers behave: For example, if a, b, c, d are four elements of $\mathbb{K}$ such that a and c are invertible, then $\dfrac{b}{a} + \dfrac{d}{c} = \dfrac{bc+da}{ac}$ and $\dfrac{b}{a} \cdot \dfrac{d}{c} = \dfrac{bd}{ac}$ (and the product ac is indeed invertible, so that the fractions $\dfrac{bc+da}{ac}$ and $\dfrac{bd}{ac}$ actually make sense).

Of course, the meaning of the word "invertible" depends on the ring $\mathbb{K}$. For example, the integer 2 is invertible in $\mathbb{Q}$ (because $\dfrac{1}{2}$ is an inverse of 2 in $\mathbb{Q}$), but not invertible in $\mathbb{Z}$ (since it has no inverse in $\mathbb{Z}$). Thus, it is important to say "invertible in $\mathbb{K}$" unless the context makes it clear what $\mathbb{K}$ is.

One can usually work with invertible elements in commutative rings in the same way as one works with nonzero rational numbers. For example, if a is an invertible element of $\mathbb{K}$, then we can define a^n not only for all $n \in \mathbb{N}$, but also for all $n \in \mathbb{Z}$ (by setting $a^n = \left(a^{-1}\right)^{-n}$ for all negative integers n). Of course, when $n = -1$, this is consistent with our notation a^{-1} for the inverse of a.

Next, we define the notion of a *field*[242].

[242]We are going to use the following simple fact: A commutative ring $\mathbb{K}$ is a trivial ring if and only if $0_{\mathbb{K}} = 1_{\mathbb{K}}$.

Definition 6.67. A commutative ring $\mathbb{K}$ is said to be a *field* if it satisfies the following two properties:

- We have $0_{\mathbb{K}} \neq 1_{\mathbb{K}}$ (that is, $\mathbb{K}$ is not a trivial ring).
- Every element of $\mathbb{K}$ is either zero or invertible.

For example, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields, whereas polynomial rings such as $\mathbb{Q}[x]$ or $\mathbb{R}[a,b]$ are not fields[243]. For n being a positive integer, the ring $\mathbb{Z}/n\mathbb{Z}$ (that is, the ring of residue classes of integers modulo n) is a field if and only if n is a prime number.

Linear algebra (i.e., the study of matrices and linear transformations) becomes much easier (in many aspects) when $\mathbb{K}$ is a field[244]. This is one of the main reasons why most courses on linear algebra work over fields only (or begin by working over fields and only later move to the generality of commutative rings). In these notes we are almost completely limiting ourselves to the parts of matrix theory which work over any commutative ring. Nevertheless, let us comment on how determinants can be computed fast when $\mathbb{K}$ is a field.

Proof. Assume that $\mathbb{K}$ is a trivial ring. Thus, $\mathbb{K}$ has only one element. Hence, both $0_{\mathbb{K}}$ and $1_{\mathbb{K}}$ have to equal this one element. Therefore, $0_{\mathbb{K}} = 1_{\mathbb{K}}$.

Now, forget that we assumed that $\mathbb{K}$ is a trivial ring. We thus have proven that

$$\text{if } \mathbb{K} \text{ is a trivial ring, then } 0_{\mathbb{K}} = 1_{\mathbb{K}}. \tag{415}$$

Conversely, assume that $0_{\mathbb{K}} = 1_{\mathbb{K}}$. Then, every $a \in \mathbb{K}$ satisfies $a = a \cdot \underbrace{1_{\mathbb{K}}}_{=0_{\mathbb{K}}} = a \cdot 0_{\mathbb{K}} = 0_{\mathbb{K}} \in \{0_{\mathbb{K}}\}$. In other words, $\mathbb{K} \subseteq \{0_{\mathbb{K}}\}$. Combining this with $\{0_{\mathbb{K}}\} \subseteq \mathbb{K}$, we obtain $\mathbb{K} = \{0_{\mathbb{K}}\}$. Hence, $\mathbb{K}$ has only one element. In other words, $\mathbb{K}$ is a trivial ring.

Now, forget that we assumed that $0_{\mathbb{K}} = 1_{\mathbb{K}}$. We thus have proven that

$$\text{if } 0_{\mathbb{K}} = 1_{\mathbb{K}}, \text{ then } \mathbb{K} \text{ is a trivial ring.}$$

Combining this with (415), we conclude that $\mathbb{K}$ is a trivial ring if and only if $0_{\mathbb{K}} = 1_{\mathbb{K}}$.

[243] For example, the polynomial x is not invertible in $\mathbb{Q}[x]$.

[244] Many properties of a matrix over a field (such as its rank) are not even well-defined over an arbitrary commutative ring.

Remark 6.68. Assume that $\mathbb{K}$ is a field. If A is an $n \times n$-matrix over $\mathbb{K}$, then the determinant of A can be computed using (340)... but in practice, you probably do not **want** to compute it this way, since the right hand side of (340) contains a sum of $n!$ terms.

It turns out that there is an algorithm to compute $\det A$, which is (usually) a lot faster. It is a version of the Gaussian elimination algorithm commonly used for solving systems of linear equations.

Let us illustrate it on an example: Set

$$n = 4, \qquad \mathbb{K} = \mathbb{Q} \qquad \text{and } A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & -1 & 0 & 2 \\ 2 & 4 & -2 & 3 \\ 5 & 1 & 3 & 5 \end{pmatrix}.$$

We want to find $\det A$.

Exercise 6.8 **(b)** shows that if we add a scalar multiple of a column of a matrix to another column of this matrix, then the determinant of the matrix does not change. Now, by adding appropriate scalar multiples of the fourth column of A to the first three columns of A, we can make sure that the first three entries of the fourth row of A become zero: Namely, we have to

- add (-1) times the fourth column of A to the first column of A;
- add $(-1/5)$ times the fourth column of A to the second column of A;
- add $(-3/5)$ times the fourth column of A to the third column of A.

These additions can be performed in any order, since none of them "interacts" with any other (more precisely, none of them uses any entries that another of them changes). As we know, none of these additions changes the determinant of the matrix.

Having performed these three additions, we end up with the matrix

$$A' = \begin{pmatrix} 1 & 2 & 3 & 0 \\ -2 & -7/5 & -6/5 & 2 \\ -1 & 17/5 & -19/5 & 3 \\ 0 & 0 & 0 & 5 \end{pmatrix}. \tag{416}$$

We have $\det(A') = \det A$ (because A' was obtained from A by three operations which do not change the determinant). Moreover, the fourth row of A' contains only one nonzero entry – namely, its last entry. In other words, if we write A' in the form $A' = \left(a'_{i,j}\right)_{1\leq i\leq 4,\ 1\leq j\leq 4}$, then $a'_{4,j} = 0$ for every $j \in \{1,2,3\}$. Thus, Theorem 6.43 (applied to 4, A' and $a'_{i,j}$ instead of n, A and $a_{i,j}$) shows that

$$\det\left(A'\right) = \underbrace{a'_{4,4}}_{=5} \cdot \det \left(\underbrace{\left(a'_{i,j}\right)_{1\leq i\leq 3,\ 1\leq j\leq 3}}_{=\begin{pmatrix} 1 & 2 & 3 \\ -2 & -7/5 & -6/5 \\ -1 & 17/5 & -19/5 \end{pmatrix}} \right)$$

$$= 5 \cdot \det \begin{pmatrix} 1 & 2 & 3 \\ -2 & -7/5 & -6/5 \\ -1 & 17/5 & -19/5 \end{pmatrix}.$$

Comparing this with $\det\left(A'\right) = \det A$, we obtain

$$\det A = 5 \cdot \det \begin{pmatrix} 1 & 2 & 3 \\ -2 & -7/5 & -6/5 \\ -1 & 17/5 & -19/5 \end{pmatrix}.$$

Thus, we have reduced the problem of computing $\det A$ (the determinant of a 4×4-matrix) to the problem of computing $\det \begin{pmatrix} 1 & 2 & 3 \\ -2 & -7/5 & -6/5 \\ -1 & 17/5 & -19/5 \end{pmatrix}$ (the determinant of a 3×3-matrix). Likewise, we can try to reduce the latter problem to the computation of the determinant of a 2×2-matrix, and then further to the computation of the determinant of a 1×1-matrix. (In our example, we obtain $\det A = -140$ at the end.)

This looks like a viable algorithm (which is, furthermore, fairly fast: essentially as fast as Gaussian elimination). But does it always work? It turns out that it **almost** always works. There are cases in which it can get "stuck", and it needs to be modified to deal with these cases.

Namely, what can happen is that the (n, n)-th entry of the matrix A could be 0. Again, let us observe this on an example: Set $n = 4$ and $A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & -1 & 0 & 2 \\ 2 & 4 & -2 & 3 \\ 5 & 1 & 3 & 0 \end{pmatrix}$. Then, we cannot turn the first three entries of the fourth row of A into zeroes by adding appropriate multiples of the fourth column to the first three columns. (Whatever multiples we add, the fourth row stays unchanged.) However, we can now switch the second row of A with the fourth row. This operation produces the matrix $B = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 5 & 1 & 3 & 0 \\ 2 & 4 & -2 & 3 \\ 0 & -1 & 0 & 2 \end{pmatrix}$, which satisfies $\det B = -\det A$ (by Exercise

6.7 **(a)**). Thus, it suffices to compute $\det B$; and this can be done as above.

The reason why we switched the second row of A with the fourth row is that the last entry of the second row of A was nonzero. In general, we need to find a $k \in \{1, 2, \ldots, n\}$ such that the last entry of the k-th row of A is nonzero, and switch the k-th row of A with the n-th row. But what if no such k exists? In this case, we need another way to compute $\det A$. It turns out that this is very easy: If there is no $k \in \{1, 2, \ldots, n\}$ such that the last entry of the k-th row of A is nonzero, then the last column of A consists of zeroes, and thus Exercise 6.7 **(d)** shows that $\det A = 0$.

When $\mathbb{K}$ is not a field, this algorithm breaks (or, at least, **can** break). Indeed, it relies on the fact that the (n, n)-th entry of the matrix A is either zero or invertible. Over a commutative ring $\mathbb{K}$, it might be neither. For example, if we had tried to work with $\mathbb{K} = \mathbb{Z}$ (instead of $\mathbb{K} = \mathbb{Q}$) in our above example, then we would not be able to add $(-1/5)$ times the fourth column of A to the second column of A (because $-1/5 \notin \mathbb{Z} = \mathbb{K}$). Fortunately, of course, $\mathbb{Z}$ is a subset of $\mathbb{Q}$ (and its operations $+$ and $\cdot$ are consistent with those of $\mathbb{Q}$), so that we can just perform the whole algorithm over $\mathbb{Q}$ instead of $\mathbb{Z}$. However, we aren't always in luck: Some commutative rings $\mathbb{K}$ cannot be "embedded" into fields in the way $\mathbb{Z}$ is embedded into $\mathbb{Q}$. (For instance, $\mathbb{Z}/4\mathbb{Z}$ cannot be embedded into a field.)

Nevertheless, there **are** reasonably fast algorithms for computing determinants over any commutative ring; see [Rote01, §2].

6.9. The Cauchy determinant

Now, we can state another classical formula for a determinant: the *Cauchy determinant*. In one of its many forms, it says the following:

Exercise 6.18. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $y_1, y_2, \ldots, y_n$ be n elements of $\mathbb{K}$. Assume that $x_i + y_j$ is invertible in $\mathbb{K}$ for every $(i, j) \in \{1, 2, \ldots, n\}^2$. Then, prove that

$$\det\left(\left(\frac{1}{x_i + y_j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = \frac{\prod\limits_{1 \leq i < j \leq n} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,n\}^2} \left(x_i + y_j\right)}.$$

There is a different version of the Cauchy determinant floating around in literature; it differs from Exercise 6.18 in that each "$x_i + y_j$" is replaced by "$x_i - y_j$", and in that "$y_i - y_j$" is replaced by "$y_j - y_i$". Of course, this version is nothing else than the result of applying Exercise 6.18 to $-y_1, -y_2, \ldots, -y_n$ instead of $y_1, y_2, \ldots, y_n$.

Exercise 6.19. Let n be a positive integer. Let $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n\times n$-matrix such that $a_{n,n}$ is invertible (in $\mathbb{K}$). Prove that

$$\det\left(\left(a_{i,j}a_{n,n}-a_{i,n}a_{n,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right) = a_{n,n}^{n-2}\cdot\det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right). \tag{417}$$

Exercise 6.19 is known as the *Chio pivotal condensation theorem*[245].

Remark 6.69. Exercise 6.19 gives a way to reduce the computation of an $n\times n$-determinant (the one on the right hand side of (417)) to the computation of an $(n-1)\times(n-1)$-determinant (the one on the left hand side), provided that $a_{n,n}$ is invertible. If this reminds you of Remark 6.68, you are thinking right...

Remark 6.70. Exercise 6.19 holds even without the assumption that $a_{n,n}$ be invertible, as long as we assume (instead) that $n\geq 2$. (If we don't assume that $n\geq 2$, then the $a_{n,n}^{n-2}$ on the right hand side of (417) will not be defined for non-invertible $a_{n,n}$.) Proving this is beyond these notes, though. (A proof of this generalized version of Exercise 6.19 can be found in [KarZha16]. It can also be obtained as a particular case of [BerBru08, (4)][246].)

6.10. Further determinant equalities

Next, let us provide an assortment of other exercises on determinants. Hundreds of exercises (ranging from easy to challenging) on the properties and evaluations of determinants can be found in [FadSom72, Chapter 2], and some more in [Prasol94, Chapter I]; in comparison, our selection is rather small.

Exercise 6.20. Let $n\in\mathbb{N}$. Let $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n\times n$-matrix. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$. Prove that

$$\sum_{k=1}^{n}\det\left(\left(a_{i,j}b_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \left(b_1+b_2+\cdots+b_n\right)\det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right),$$

where $\delta_{j,k}$ means the nonnegative integer $\begin{cases}1, & \text{if } j=k;\\ 0, & \text{if } j\neq k\end{cases}$. Equivalently (in more

[245] See [Heinig11, footnote 2] and [Abeles14, §2] for some hints about its history. A variant of the formula (singling out the 1-st row and the 1-st column instead of the n-th row and the n-th column) appears in [Heffer17, Chapter Four, Topic "Chio's Method"].

[246] In more detail: If we apply [BerBru08, (4)] to $k=n-1$, then the right hand side is precisely $\det\left(\left(a_{i,j}a_{n,n}-a_{i,n}a_{n,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)$, and so the formula becomes (417).

reader-friendly terms): Prove that

$$\det\begin{pmatrix} a_{1,1}b_1 & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1}b_2 & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}b_n & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} + \det\begin{pmatrix} a_{1,1} & a_{1,2}b_1 & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2}b_2 & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2}b_n & \cdots & a_{n,n} \end{pmatrix}$$
$$+ \cdots + \det\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n}b_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n}b_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n}b_n \end{pmatrix}$$
$$= (b_1 + b_2 + \cdots + b_n)\det\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix}.$$

Exercise 6.21. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $x \in \mathbb{K}$. Prove that

$$\det\begin{pmatrix} x & a_1 & a_2 & \cdots & a_{n-1} & a_n \\ a_1 & x & a_2 & \cdots & a_{n-1} & a_n \\ a_1 & a_2 & x & \cdots & a_{n-1} & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 & a_2 & a_3 & \cdots & x & a_n \\ a_1 & a_2 & a_3 & \cdots & a_n & x \end{pmatrix} = \left(x + \sum_{i=1}^{n} a_i\right) \prod_{i=1}^{n} (x - a_i).$$

Exercise 6.22. Let $n > 1$ be an integer. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$. Let A be the $n \times n$-matrix

$$\left(\begin{cases} a_j, & \text{if } i = j; \\ b_j, & \text{if } i \equiv j + 1 \operatorname{mod} n; \\ 0, & \text{otherwise} \end{cases}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$$
$$= \begin{pmatrix} a_1 & 0 & 0 & \cdots & 0 & b_n \\ b_1 & a_2 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & b_{n-1} & a_n \end{pmatrix}.$$

Prove that

$$\det A = a_1 a_2 \cdots a_n + (-1)^{n-1} b_1 b_2 \cdots b_n.$$

Remark 6.71. If we replace "$i \equiv j + 1 \mod n$" by "$i \equiv j + 2 \mod n$" in Exercise 6.22, then the pattern can break. For instance, for $n = 4$ we have

$$\det \begin{pmatrix} a_1 & 0 & b_3 & 0 \\ 0 & a_2 & 0 & b_4 \\ b_1 & 0 & a_3 & 0 \\ 0 & b_2 & 0 & a_4 \end{pmatrix} = (a_2a_4 - b_2b_4)(a_1a_3 - b_1b_3),$$

which is not of the form $a_1a_2a_3a_4 \pm b_1b_2b_3b_4$ anymore. Can you guess for which $d \in \{1, 2, \ldots, n-1\}$ we can replace "$i \equiv j + 1 \mod n$" by "$i \equiv j + d \mod n$" in Exercise 6.22 and still get a formula of the form $\det A = a_1a_2 \cdots a_n \pm b_1b_2 \cdots b_n$? (The answer to this question requires a little bit of elementary number theory – namely, the concept of "coprimality".)

6.11. Alternating matrices

Our next two exercises will concern two special classes of matrices: the *antisymmetric* and the *alternating matrices*. Let us first define these classes:

Definition 6.72. Let $n \in \mathbb{N}$. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix.

(a) The matrix A is said to be *antisymmetric* if and only if $A^T = -A$. (Recall that A^T is defined as in Definition 6.10.)

(b) The matrix A is said to be *alternating* if and only if it satisfies $A^T = -A$ and ($a_{i,i} = 0$ for all $i \in \{1, 2, \ldots, n\}$).

Example 6.73. A 1×1-matrix is alternating if and only if it is the zero matrix $0_{1\times 1} = \begin{pmatrix} 0 \end{pmatrix}$.

A 2×2-matrix is alternating if and only if it has the form $\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix}$ for some $a \in \mathbb{K}$.

A 3×3-matrix is alternating if and only if it has the form $\begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$ for some $a, b, c \in \mathbb{K}$.

Visually speaking, an $n \times n$-matrix is alternating if and only if its diagonal entries are 0 and its entries below the diagonal are the negatives of their "mirror-image" entries above the diagonal.

Remark 6.74. Clearly, any alternating matrix is antisymmetric. It is easy to see that an $n \times n$-matrix $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ is antisymmetric if and only if every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfies $a_{i,j} = -a_{j,i}$. Thus, if $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ is antisymmetric, then every $i \in \{1, 2, \ldots, n\}$ satisfies $a_{i,i} = -a_{i,i}$ and thus $2a_{i,i} = 0$. If $\mathbb{K}$ is one of the rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, then we can cancel 2 from this last equality, and conclude that every antisymmetric $n \times n$-matrix A is alternating. However, there are commutative rings $\mathbb{K}$ for which this does not hold (for example, the ring $\mathbb{Z}/2\mathbb{Z}$ of integers modulo 2).

Antisymmetric matrices are also known as *skew-symmetric* matrices.

Exercise 6.23. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an alternating $n \times n$-matrix. Let S be an $n \times m$-matrix. Prove that the $m \times m$-matrix $S^T AS$ is alternating.

Exercise 6.24. Let $n \in \mathbb{N}$ be odd. Let A be an $n \times n$-matrix. Prove the following:

(a) If A is antisymmetric, then $2 \det A = 0$.

(b) If A is alternating, then $\det A = 0$.

Remark 6.75. If $\mathbb{K}$ is one of the rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, then Exercise 6.24 **(b)** follows from Exercise 6.24 **(a)** (because any alternating matrix is antisymmetric, and because we can cancel 2 from the equality $2 \det A = 0$). However, this quick way of solving Exercise 6.24 **(b)** does not work for general $\mathbb{K}$.

Remark 6.76. Exercise 6.24 **(b)** provides a really simple formula for $\det A$ when A is an alternating $n \times n$-matrix for **odd** n. One might wonder what can be said about $\det A$ when A is an alternating $n \times n$-matrix for **even** n. The answer is far less simple, but more interesting: It turns that $\det A$ is the square of a certain element of $\mathbb{K}$, called the *Pfaffian* of A. See [Conrad2, (5.5)] for a short introduction into the Pfaffian (although at a less elementary level than these notes); see [BruRys91, §9.5] and [Loehr11, §12.12][247] for a more combinatorial treatment of the Pfaffian (and an application to matchings of graphs!). For example, the Pfaffian of an alternating 4×4-matrix $A = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix}$ is $af - be + cd$, and it is indeed easy to check that this matrix satisfies $\det A = (af - be + cd)^2$.

6.12. Laplace expansion

We shall now state Laplace expansion in full. We begin with an example:

Example 6.77. Let $A = (a_{i,j})_{1 \leq i \leq 3,\ 1 \leq j \leq 3}$ be a 3×3-matrix. From (343), we obtain

$$\det A = a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1}. \tag{418}$$

[247] Beware that Loehr, in [Loehr11, §12.12], seems to work only in the setting where 2 is cancellable in the ring $\mathbb{K}$ (that is, where $2a = 0$ for an element $a \in \mathbb{K}$ implies $a = 0$). Thus, Loehr does not have to distinguish between antisymmetric and alternating matrices (he calls them "skew-symmetric matrices" instead). His arguments, however, can easily be adapted to the general case.

On the right hand side of this equality, we have six terms, each of which contains either $a_{2,1}$ or $a_{2,2}$ or $a_{2,3}$. Let us combine the two terms containing $a_{2,1}$ and factor out $a_{2,1}$, then do the same with the two terms containing $a_{2,2}$, and with the two terms containing $a_{2,3}$. As a result, (418) becomes

$$\begin{aligned}
&\det A \\
&= a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1} \\
&= a_{2,1} \underbrace{\left(a_{1,3}a_{3,2} - a_{1,2}a_{3,3}\right)}_{=\det\begin{pmatrix} a_{1,3} & a_{1,2} \\ a_{3,3} & a_{3,2} \end{pmatrix}} + a_{2,2} \underbrace{\left(a_{1,1}a_{3,3} - a_{1,3}a_{3,1}\right)}_{=\det\begin{pmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{pmatrix}} + a_{2,3} \underbrace{\left(a_{1,2}a_{3,1} - a_{1,1}a_{3,2}\right)}_{=\det\begin{pmatrix} a_{1,2} & a_{1,1} \\ a_{3,2} & a_{3,1} \end{pmatrix}} \\
&= a_{2,1} \det \begin{pmatrix} a_{1,3} & a_{1,2} \\ a_{3,3} & a_{3,2} \end{pmatrix} + a_{2,2} \det \begin{pmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{pmatrix} + a_{2,3} \det \begin{pmatrix} a_{1,2} & a_{1,1} \\ a_{3,2} & a_{3,1} \end{pmatrix}. \qquad (419)
\end{aligned}$$

This is a nice formula with an obvious pattern: The right hand side can be rewritten as $\sum\limits_{q=1}^{3} a_{2,q} \det\left(B_{2,q}\right)$, where $B_{2,q} = \begin{pmatrix} a_{1,q+2} & a_{1,q+1} \\ a_{3,q+2} & a_{3,q+1} \end{pmatrix}$ (where we set $a_{i,4} = a_{i,1}$ and $a_{i,5} = a_{i,2}$ for all $i \in \{1,2,3\}$). Notice the cyclic symmetry (with respect to the index of the column) in this formula! Unfortunately, in this exact form, the formula does not generalize to bigger matrices (or even to smaller: the analogue for a 2×2-matrix would be $\det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = -a_{2,1}a_{1,2} + a_{2,2}a_{1,1}$, which has a minus sign unlike $\sum\limits_{q=1}^{3} a_{2,q} \det\left(B_{2,q}\right)$).

However, we can slightly modify our formula, sacrificing the cyclic symmetry but making it generalize. Namely, let us rewrite $a_{1,3}a_{3,2} - a_{1,2}a_{3,3}$ as $-\left(a_{1,2}a_{3,3} - a_{1,3}a_{3,2}\right)$ and $a_{1,2}a_{3,1} - a_{1,1}a_{3,2}$ as $-\left(a_{1,1}a_{3,2} - a_{1,2}a_{3,1}\right)$; we thus obtain

$$\begin{aligned}
&\det A \\
&= a_{2,1} \underbrace{\left(a_{1,3}a_{3,2} - a_{1,2}a_{3,3}\right)}_{=-\left(a_{1,2}a_{3,3} - a_{1,3}a_{3,2}\right)} + a_{2,2} \left(a_{1,1}a_{3,3} - a_{1,3}a_{3,1}\right) + a_{2,3} \underbrace{\left(a_{1,2}a_{3,1} - a_{1,1}a_{3,2}\right)}_{=-\left(a_{1,1}a_{3,2} - a_{1,2}a_{3,1}\right)} \\
&= -a_{2,1} \underbrace{\left(a_{1,2}a_{3,3} - a_{1,3}a_{3,2}\right)}_{=\det\begin{pmatrix} a_{1,2} & a_{1,3} \\ a_{3,2} & a_{3,3} \end{pmatrix}} + a_{2,2} \underbrace{\left(a_{1,1}a_{3,3} - a_{1,3}a_{3,1}\right)}_{=\det\begin{pmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{pmatrix}} - a_{2,3} \underbrace{\left(a_{1,1}a_{3,2} - a_{1,2}a_{3,1}\right)}_{=\det\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{3,1} & a_{3,2} \end{pmatrix}} \\
&= -a_{2,1} \det \begin{pmatrix} a_{1,2} & a_{1,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} + a_{2,2} \det \begin{pmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{pmatrix} - a_{2,3} \det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{3,1} & a_{3,2} \end{pmatrix} \\
&= \sum_{q=1}^{3} (-1)^q a_{2,q} \det\left(C_{2,q}\right), \qquad (420)
\end{aligned}$$

where $C_{2,q}$ means the matrix obtained from A by crossing out the 2-nd row and the q-th column. This formula (unlike (419)) involves powers of -1, but it can be generalized.

How? First, we notice that we can find a similar formula by factoring out $a_{1,1}, a_{1,2}, a_{1,3}$ (instead of $a_{2,1}, a_{2,2}, a_{2,3}$); this formula will be

$$\det A = \sum_{q=1}^{3} (-1)^{q-1} a_{1,q} \det \left(C_{1,q} \right),$$

where $C_{1,q}$ means the matrix obtained from A by crossing out the 1-st row and the q-th column. This formula, and (420), suggest the following generalization: If $A = \left(a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ is an $n \times n$-matrix, and if $p \in \{1, 2, \ldots, n\}$, then

$$\det A = \sum_{q=1}^{n} (-1)^{p+q} a_{p,q} \det \left(C_{p,q} \right), \tag{421}$$

where $C_{p,q}$ means the matrix obtained from A by crossing out the p-th row and the q-th column. (The only part of this formula which is not easy to guess is $(-1)^{p+q}$; you might need to compute several particular cases to guess this pattern. Of course, you could also have guessed $(-1)^{p-q}$ or $(-1)^{q-p}$ instead, because $(-1)^{p+q} = (-1)^{p-q} = (-1)^{q-p}$.)

The formula (421) is what is usually called the Laplace expansion with respect to the p-th row. We will prove it below (Theorem 6.82 **(a)**), and we will also prove an analogous "Laplace expansion with respect to the q-th column" (Theorem 6.82 **(b)**).

Let us first define a notation:

Definition 6.78. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = \left(a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$ be an $n \times m$-matrix. Let $i_1, i_2, \ldots, i_u$ be some elements of $\{1, 2, \ldots, n\}$; let $j_1, j_2, \ldots, j_v$ be some elements of $\{1, 2, \ldots, m\}$. Then, we define $\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} A$ to be the $u \times v$-matrix $\left(a_{i_x, j_y} \right)_{1 \leq x \leq u,\ 1 \leq y \leq v}$.

When $i_1 < i_2 < \cdots < i_u$ and $j_1 < j_2 < \cdots < j_v$, the matrix $\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} A$ can be obtained from A by crossing out all rows other than the i_1-th, the i_2-th, etc., the i_u-th row and crossing out all columns other than the j_1-th, the j_2-th, etc., the j_v-th column. Thus, in this case, $\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} A$ is called a *submatrix* of A.

For example, if $n = 3$, $m = 4$ and $A = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & \ell \end{pmatrix}$, then $\operatorname{sub}_{1,3}^{2,3,4} A = \begin{pmatrix} b & c & d \\ j & k & \ell \end{pmatrix}$ (this is a submatrix of A) and $\operatorname{sub}_{2,3}^{3,1,1} A = \begin{pmatrix} g & e & e \\ k & i & i \end{pmatrix}$ (this is not, in general, a submatrix of A).

The following properties follow trivially from the definitions:

Proposition 6.79. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Recall the notations introduced in Definition 6.31.

(a) We have $\operatorname{sub}_{1,2,\ldots,n}^{1,2,\ldots,m} A = A$.

(b) If $i_1, i_2, \ldots, i_u$ are some elements of $\{1, 2, \ldots, n\}$, then

$$\operatorname{rows}_{i_1,i_2,\ldots,i_u} A = \operatorname{sub}_{i_1,i_2,\ldots,i_u}^{1,2,\ldots,m} A.$$

(c) If $j_1, j_2, \ldots, j_v$ are some elements of $\{1, 2, \ldots, m\}$, then

$$\operatorname{cols}_{j_1,j_2,\ldots,j_v} A = \operatorname{sub}_{1,2,\ldots,n}^{j_1,j_2,\ldots,j_v} A.$$

(d) Let $i_1, i_2, \ldots, i_u$ be some elements of $\{1, 2, \ldots, n\}$; let $j_1, j_2, \ldots, j_v$ be some elements of $\{1, 2, \ldots, m\}$. Then,

$$\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v} A = \operatorname{rows}_{i_1,i_2,\ldots,i_u} \left(\operatorname{cols}_{j_1,j_2,\ldots,j_v} A\right) = \operatorname{cols}_{j_1,j_2,\ldots,j_v} \left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right).$$

(e) Let $i_1, i_2, \ldots, i_u$ be some elements of $\{1, 2, \ldots, n\}$; let $j_1, j_2, \ldots, j_v$ be some elements of $\{1, 2, \ldots, m\}$. Then,

$$\left(\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v} A\right)^T = \operatorname{sub}_{j_1,j_2,\ldots,j_v}^{i_1,i_2,\ldots,i_u} \left(A^T\right).$$

Definition 6.80. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n objects. Let $i \in \{1, 2, \ldots, n\}$. Then, $(a_1, a_2, \ldots, \widehat{a_i}, \ldots, a_n)$ shall mean the list $(a_1, a_2, \ldots, a_{i-1}, a_{i+1}, a_{i+2}, \ldots, a_n)$ (that is, the list $(a_1, a_2, \ldots, a_n)$ with its i-th entry removed). (Thus, the "hat" over the a_i means that this a_i is being omitted from the list.)

For example, $\left(1^2, 2^2, \ldots, \widehat{5^2}, \ldots, 8^2\right) = \left(1^2, 2^2, 3^2, 4^2, 6^2, 7^2, 8^2\right)$.

Definition 6.81. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$, we let $A_{\sim i, \sim j}$ be the $(n-1) \times (m-1)$-matrix $\operatorname{sub}_{1,2,\ldots,\widehat{i},\ldots,n}^{1,2,\ldots,\widehat{j},\ldots,m} A$. (Thus, $A_{\sim i, \sim j}$ is the matrix obtained from A by crossing out the i-th row and the j-th column.)

For example, if $n = m = 3$ and $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, then $A_{\sim 1, \sim 2} = \begin{pmatrix} d & f \\ g & i \end{pmatrix}$ and $A_{\sim 3, \sim 2} = \begin{pmatrix} a & c \\ d & f \end{pmatrix}$.

The notation $A_{\sim i, \sim j}$ introduced in Definition 6.81 is not very standard; but there does not seem to be a standard one[248].

Now we can finally state Laplace expansion:

[248] For example, Gill Williamson uses the notation $A(i \mid j)$ in [Willia18, Chapter 3].

Theorem 6.82. Let $n \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix.

(a) For every $p \in \{1, 2, \ldots, n\}$, we have

$$\det A = \sum_{q=1}^{n} (-1)^{p+q} a_{p,q} \det \left(A_{\sim p, \sim q}\right).$$

(b) For every $q \in \{1, 2, \ldots, n\}$, we have

$$\det A = \sum_{p=1}^{n} (-1)^{p+q} a_{p,q} \det \left(A_{\sim p, \sim q}\right).$$

Theorem 6.82 **(a)** is known as the *Laplace expansion along the p-th row* (or *Laplace expansion with respect to the p-th row*), whereas Theorem 6.82 **(b)** is known as the *Laplace expansion along the q-th column* (or *Laplace expansion with respect to the q-th column*). Notice that Theorem 6.82 **(a)** is equivalent to the formula (421), because the $A_{\sim p, \sim q}$ in Theorem 6.82 **(a)** is precisely what we called $C_{p,q}$ in (421).

We prepare the field for the proof of Theorem 6.82 with a few lemmas.

Lemma 6.83. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $n \in \mathbb{N}$. For every $p \in [n]$, we define a permutation $g_p \in S_n$ by $g_p = \operatorname{cyc}_{p, p+1, \ldots, n}$ (where we are using the notations of Definition 5.37).

(a) We have $\left(g_p(1), g_p(2), \ldots, g_p(n-1)\right) = (1, 2, \ldots, \widehat{p}, \ldots, n)$ for every $p \in [n]$.

(b) We have $(-1)^{g_p} = (-1)^{n-p}$ for every $p \in [n]$.

(c) Let $p \in [n]$. We define a map

$$g'_p : [n-1] \to [n] \setminus \{p\}$$

by

$$\left(g'_p(i) = g_p(i) \qquad \text{for every } i \in [n-1]\right).$$

This map g'_p is well-defined and bijective.

(d) Let $p \in [n]$ and $q \in [n]$. We define a map

$$T : \{\tau \in S_n \mid \tau(n) = n\} \to \{\tau \in S_n \mid \tau(p) = q\}$$

by

$$\left(T(\sigma) = g_q \circ \sigma \circ \left(g_p\right)^{-1} \qquad \text{for every } \sigma \in \{\tau \in S_n \mid \tau(n) = n\}\right).$$

Then, this map T is well-defined and bijective.

Proof of Lemma 6.83. **(a)** This is trivial.

(b) Let $p \in [n]$. Exercise 5.17 **(d)** (applied to $k = p+1$ and $(i_1, i_2, \ldots, i_k) = (p, p+1, \ldots, n)$) yields

$$(-1)^{\operatorname{cyc}_{p,p+1,\ldots,n}} = (-1)^{n-(p+1)-1} = (-1)^{n-p-2} = (-1)^{n-p}.$$

Now, $g_p = \operatorname{cyc}_{p,p+1,\ldots,n}$, so that $(-1)^{g_p} = (-1)^{\operatorname{cyc}_{p,p+1,\ldots,n}} = (-1)^{n-p}$. This proves Lemma 6.83 **(b)**.

(c) We have $g_p(n) = p$ (since $g_p = \operatorname{cyc}_{p,p+1,\ldots,n}$). Also, g_p is injective (since g_p is a permutation). Therefore, for every $i \in [n-1]$, we have

$$g_p(i) \neq g_p(n) \qquad \left(\text{since } i \neq n \text{ (because } i \in [n-1]\text{) and since } g_p \text{ is injective}\right)$$
$$= p,$$

so that $g_p(i) \in [n] \setminus \{p\}$. This shows that the map g'_p is well-defined.

To prove that g'_p is bijective, we can construct its inverse. Indeed, for every $i \in [n] \setminus \{p\}$, we have

$$\left(g_p\right)^{-1}(i) \neq n \qquad \left(\text{since } i \neq p = g_p(n)\right)$$

and thus $\left(g_p\right)^{-1}(i) \in [n-1]$. Hence, we can define a map $h : [n] \setminus \{p\} \to [n-1]$ by

$$\left(h(i) = \left(g_p\right)^{-1}(i) \qquad \text{for every } i \in [n] \setminus \{p\}\right).$$

It is straightforward to check that the maps g'_p and h are mutually inverse. Thus, g'_p is bijective. Lemma 6.83 **(c)** is thus proven.

(d) We have $g_p(n) = p$ (since $g_p = \operatorname{cyc}_{p,p+1,\ldots,n}$) and $g_q(n) = q$ (similarly). Hence, $\left(g_p\right)^{-1}(p) = n$ (since $g_p(n) = p$) and $\left(g_q\right)^{-1}(q) = n$ (since $g_q(n) = q$).

For every $\sigma \in \{\tau \in S_n \mid \tau(n) = n\}$, we have $\sigma(n) = n$ and thus

$$\left(g_q \circ \sigma \circ \left(g_p\right)^{-1}\right)(p) = g_q\left(\sigma\left(\underbrace{\left(g_p\right)^{-1}(p)}_{=n}\right)\right) = g_q\left(\underbrace{\sigma(n)}_{=n}\right) = g_q(n) = q$$

and therefore $g_q \circ \sigma \circ \left(g_p\right)^{-1} \in \{\tau \in S_n \mid \tau(p) = q\}$. Thus, the map T is well-defined.

We can also define a map

$$Q : \{\tau \in S_n \mid \tau(p) = q\} \to \{\tau \in S_n \mid \tau(n) = n\}$$

by

$$\left(Q(\sigma) = \left(g_q\right)^{-1} \circ \sigma \circ g_p \qquad \text{for every } \sigma \in \{\tau \in S_n \mid \tau(p) = q\}\right).$$

The well-definedness of Q can be checked similarly to how we proved the well-definedness of T. It is straightforward to verify that the maps Q and T are mutually inverse. Thus, T is bijective. This completes the proof of Lemma 6.83 **(d)**. □

Our next step towards the proof of Theorem 6.82 is the following lemma:

Lemma 6.84. Let $n \in \mathbb{N}$. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. Let $p \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, n\}$. Then,

$$\sum_{\substack{\sigma \in S_n; \\ \sigma(p) = q}} (-1)^{\sigma} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)} = (-1)^{p+q} \det (A_{\sim p, \sim q}) .$$

Proof of Lemma 6.84. Let us use all notations introduced in Lemma 6.83.

We have $p \in \{1, 2, \ldots, n\} = [n]$. Hence, g_p is well-defined. Similarly, g_q is well-defined. We have

$$(g_p(1), g_p(2), \ldots, g_p(n-1)) = (1, 2, \ldots, \widehat{p}, \ldots, n) \tag{422}$$

(by Lemma 6.83 **(a)**) and

$$(g_q(1), g_q(2), \ldots, g_q(n-1)) = (1, 2, \ldots, \widehat{q}, \ldots, n) \tag{423}$$

(by Lemma 6.83 **(a)**, applied to q instead of p). Now, the definition of $A_{\sim p, \sim q}$ yields

$$\begin{aligned}
A_{\sim p, \sim q} &= \operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,n} A = \operatorname{sub}^{g_q(1), g_q(2), \ldots, g_q(n-1)}_{g_p(1), g_p(2), \ldots, g_p(n-1)} A \qquad \text{(by (422) and (423))} \\
&= \left(a_{g_p(x), g_q(y)}\right)_{1 \leq x \leq n-1,\ 1 \leq y \leq n-1} \\
&\qquad \left(\text{by the definition of } \operatorname{sub}^{g_q(1), g_q(2), \ldots, g_q(n-1)}_{g_p(1), g_p(2), \ldots, g_p(n-1)} A\right) \\
&= \left(a_{g_p(i), g_q(j)}\right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \\
&\qquad (\text{here, we renamed the index } (x, y) \text{ as } (i, j)) .
\end{aligned} \tag{424}$$

Also, $[n]$ is nonempty (since $p \in [n]$), and thus we have $n > 0$.

Now, let us recall the map $T : \{\tau \in S_n \mid \tau(n) = n\} \to \{\tau \in S_n \mid \tau(p) = q\}$ defined in Lemma 6.83 **(d)**. Lemma 6.83 **(d)** says that this map T is well-defined and bijective. Every $\sigma \in \{\tau \in S_n \mid \tau(n) = n\}$ satisfies

$$(-1)^{T(\sigma)} = (-1)^{p+q} \cdot (-1)^{\sigma} \tag{425}$$

[249] and

$$\prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,(T(\sigma))(i)} = \prod_{i=1}^{n-1} a_{g_p(i), g_q(\sigma(i))} \tag{426}$$

[249] *Proof of (425):* Let $\sigma \in \{\tau \in S_n \mid \tau(n) = n\}$. Applying Lemma 6.83 **(b)** to q instead of p, we obtain $(-1)^{g_q} = (-1)^{n-q} = (-1)^{n+q}$ (since $n - q \equiv n + q \operatorname{mod} 2$).

The definition of $T(\sigma)$ yields $T(\sigma) = g_q \circ \sigma \circ (g_p)^{-1}$. Thus,

$$\underbrace{T(\sigma)}_{= g_q \circ \sigma \circ (g_p)^{-1}} \circ g_p = g_q \circ \sigma \circ \underbrace{(g_p)^{-1} \circ g_p}_{= \mathrm{id}} = g_q \circ \sigma,$$

[250].

so that

$$(-1)^{T(\sigma)\circ g_p} = (-1)^{g_q\circ\sigma} = \underbrace{(-1)^{g_q}}_{=(-1)^{n+q}} \cdot (-1)^{\sigma} \qquad \left(\text{by (315), applied to } g_q \text{ and } \sigma \text{ instead of } \sigma \text{ and } \tau\right)$$
$$= (-1)^{n+q} \cdot (-1)^{\sigma}.$$

Compared with

$$(-1)^{T(\sigma)\circ g_p} = (-1)^{T(\sigma)} \cdot \underbrace{(-1)^{g_p}}_{\substack{=(-1)^{n-p} \\ \text{(by Lemma 6.83 (b))}}} \qquad \left(\text{by (315), applied to } T(\sigma) \text{ and } g_p \text{ instead of } \sigma \text{ and } \tau\right)$$
$$= (-1)^{T(\sigma)} \cdot (-1)^{n-p},$$

this yields

$$(-1)^{T(\sigma)} \cdot (-1)^{n-p} = (-1)^{n+q} \cdot (-1)^{\sigma}.$$

We can divide both sides of this equality by $(-1)^{n-p}$ (since $(-1)^{n-p} \in \{1, -1\}$ is clearly an invertible integer), and thus we obtain

$$(-1)^{T(\sigma)} = \frac{(-1)^{n+q} \cdot (-1)^{\sigma}}{(-1)^{n-p}} = \underbrace{\frac{(-1)^{n+q}}{(-1)^{n-p}}}_{\substack{=(-1)^{(n+q)-(n-p)}=(-1)^{p+q} \\ \text{(since } (n+q)-(n-p)=p+q\text{)}}} \cdot (-1)^{\sigma} = (-1)^{p+q} \cdot (-1)^{\sigma}.$$

This proves (425).

[250]*Proof of (426):* Let $\sigma \in \{\tau \in S_n \mid \tau(n) = n\}$. Let us recall the map $g'_p : [n-1] \to [n] \setminus \{p\}$ introduced in Lemma 6.83 **(c)**. Lemma 6.83 **(c)** says that this map g'_p is well-defined and bijective. In other words, g'_p is a bijection.

Let $i \in [n-1]$. Then, $g'_p(i) = g_p(i)$ (by the definition of g'_p). Also, the definition of T yields $T(\sigma) = g_q \circ \sigma \circ (g_p)^{-1}$, so that

$$\left(\underbrace{T(\sigma)}_{=g_q\circ\sigma\circ(g_p)^{-1}}\right)\left(\underbrace{g'_p(i)}_{=g_p(i)}\right) = \left(g_q \circ \sigma \circ (g_p)^{-1}\right)(g_p(i)) = g_q\left(\sigma\left(\underbrace{(g_p)^{-1}(g_p(i))}_{=i}\right)\right) = g_q(\sigma(i)).$$

From $g'_p(i) = g_p(i)$ and $(T(\sigma))\left(g'_p(i)\right) = g_q(\sigma(i))$, we obtain

$$a_{g'_p(i),(T(\sigma))\left(g'_p(i)\right)} = a_{g_p(i),g_q(\sigma(i))}. \tag{427}$$

Now, let us forget that we fixed i. We thus have proven (427) for every $i \in [n-1]$. But now,

we have

$$\underbrace{\prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}}}_{\substack{=\prod_{\substack{i\in[n];\\ i\neq p}}\\ \text{(since }\{1,2,\ldots,n\}=[n])}} a_{i,(T(\sigma))(i)}$$

$$= \underbrace{\prod_{\substack{i\in[n];\\ i\neq p}}}_{=\prod_{i\in[n]\setminus\{p\}}} a_{i,(T(\sigma))(i)} = \prod_{i\in[n]\setminus\{p\}} a_{i,(T(\sigma))(i)} = \underbrace{\prod_{i\in[n-1]}}_{=\prod_{i=1}^{n-1}} \underbrace{a_{g'_p(i),(T(\sigma))\left(g'_p(i)\right)}}_{\substack{=a_{g_p(i),g_q(\sigma(i))}\\ \text{(by (427))}}}$$

$$\left(\begin{array}{c}\text{here, we have substituted } g'_p(i) \text{ for } i\text{, since}\\ g'_p : [n-1] \to [n]\setminus\{p\} \text{ is a bijection}\end{array}\right)$$

$$= \prod_{i=1}^{n-1} a_{g_p(i),g_q(\sigma(i))}.$$

This proves (426).

Now,

$$\underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(p)=q}}}_{=\sum_{\sigma\in\{\tau\in S_n \mid \tau(p)=q\}}} (-1)^{\sigma} \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} a_{i,\sigma(i)}$$

$$= \sum_{\sigma\in\{\tau\in S_n \mid \tau(p)=q\}} (-1)^{\sigma} \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} a_{i,\sigma(i)}$$

$$= \underbrace{\sum_{\sigma\in\{\tau\in S_n \mid \tau(n)=n\}}}_{=\sum_{\substack{\sigma\in S_n;\\ \sigma(n)=n}}} \underbrace{(-1)^{T(\sigma)}}_{\substack{=(-1)^{p+q}\cdot(-1)^{\sigma}\\ \text{(by (425))}}} \underbrace{\prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} a_{i,(T(\sigma))(i)}}_{\substack{=\prod_{i=1}^{n-1} a_{g_p(i),g_q(\sigma(i))}\\ \text{(by (426))}}}$$

$$\left(\begin{array}{c}\text{here, we have substituted } T(\sigma) \text{ for } \sigma \text{ in the sum,}\\ \text{since the map } T : \{\tau \in S_n \mid \tau(n)=n\} \to \{\tau \in S_n \mid \tau(p)=q\}\\ \text{is a bijection}\end{array}\right)$$

$$= \sum_{\substack{\sigma\in S_n;\\ \sigma(n)=n}} (-1)^{p+q} \cdot (-1)^{\sigma} \prod_{i=1}^{n-1} a_{g_p(i),g_q(\sigma(i))}$$

$$= (-1)^{p+q} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(n)=n}} (-1)^{\sigma} \prod_{i=1}^{n-1} a_{g_p(i),g_q(\sigma(i))}}_{\substack{=\det\left(\left(a_{g_p(i),g_q(j)}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)\\ \text{(by Lemma 6.44, applied to } a_{g_p(i),g_q(j)} \text{ instead of } a_{i,j})}}$$

$$= (-1)^{p+q} \det\left(\underbrace{\left(a_{g_p(i),g_q(j)}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}}_{\substack{=A_{\sim p,\sim q}\\ \text{(by (424))}}}\right) = (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right).$$

This proves Lemma 6.84. □

Now, we can finally prove Theorem 6.82:

Proof of Theorem 6.82. **(a)** Let $p \in \{1,2,\ldots,n\}$. From (341), we obtain

$$\begin{aligned}
\det A &= \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} \\
&= \sum_{q \in \{1,2,\ldots,n\}} \sum_{\substack{\sigma \in S_n; \\ \sigma(p)=q}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n}}_{= \prod_{i \in \{1,2,\ldots,n\}}} a_{i,\sigma(i)} \\
&\qquad \left(\begin{array}{c} \text{because for every } \sigma \in S_n \text{, there exists} \\ \text{exactly one } q \in \{1,2,\ldots,n\} \text{ satisfying } \sigma(p) = q \end{array} \right) \\
&= \sum_{q \in \{1,2,\ldots,n\}} \sum_{\substack{\sigma \in S_n; \\ \sigma(p)=q}} (-1)^{\sigma} \underbrace{\prod_{i \in \{1,2,\ldots,n\}} a_{i,\sigma(i)}}_{\substack{= a_{p,\sigma(p)} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)} \\ \text{(here, we have split off the factor for } i=p \text{ from the product)}}} \\
&= \underbrace{\sum_{q \in \{1,2,\ldots,n\}}}_{= \sum_{q=1}^{n}} \sum_{\substack{\sigma \in S_n; \\ \sigma(p)=q}} (-1)^{\sigma} \underbrace{a_{p,\sigma(p)}}_{\substack{= a_{p,q} \\ \text{(since } \sigma(p)=q\text{)}}} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)} \\
&= \sum_{q=1}^{n} \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma(p)=q}} (-1)^{\sigma} a_{p,q} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)}}_{= a_{p,q} \sum_{\substack{\sigma \in S_n; \\ \sigma(p)=q}} (-1)^{\sigma} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)}} \\
&= \sum_{q=1}^{n} a_{p,q} \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma(p)=q}} (-1)^{\sigma} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)}}_{\substack{= (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right) \\ \text{(by Lemma 6.84)}}} \\
&= \sum_{q=1}^{n} \underbrace{a_{p,q} (-1)^{p+q}}_{= (-1)^{p+q} a_{p,q}} \det\left(A_{\sim p,\sim q}\right) = \sum_{q=1}^{n} (-1)^{p+q} a_{p,q} \det\left(A_{\sim p,\sim q}\right).
\end{aligned}$$

This proves Theorem 6.82 **(a)**.

(b) Let $q \in \{1, 2, \ldots, n\}$. From (341), we obtain

$$\begin{aligned}
\det A &= \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} \\
&= \sum_{p \in \{1,2,\ldots,n\}} \sum_{\substack{\sigma \in S_n; \\ \sigma^{-1}(q) = p}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n}}_{= \prod_{i \in \{1,2,\ldots,n\}}} a_{i,\sigma(i)} \\
&\qquad \left(\begin{array}{c} \text{because for every } \sigma \in S_n \text{, there exists} \\ \text{exactly one } p \in \{1,2,\ldots,n\} \text{ satisfying } \sigma^{-1}(q) = p \end{array} \right) \\
&= \sum_{p \in \{1,2,\ldots,n\}} \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma^{-1}(q) = p}}}_{\substack{= \sum_{\sigma \in S_n; \ \sigma(p) = q} \\ \text{(because for any } \sigma \in S_n \text{,} \\ \text{the statement } (\sigma^{-1}(q) = p) \\ \text{is equivalent to the} \\ \text{statement } (\sigma(p) = q))}} (-1)^{\sigma} \underbrace{\prod_{i \in \{1,2,\ldots,n\}} a_{i,\sigma(i)}}_{\substack{= a_{p,\sigma(p)} \prod_{i \in \{1,2,\ldots,n\}; \ i \neq p} a_{i,\sigma(i)} \\ \text{(here, we have split off the} \\ \text{factor for } i = p \text{ from the product)}}} \\
&= \underbrace{\sum_{p \in \{1,2,\ldots,n\}}}_{= \sum_{p=1}^{n}} \sum_{\substack{\sigma \in S_n; \\ \sigma(p) = q}} (-1)^{\sigma} \underbrace{a_{p,\sigma(p)}}_{\substack{= a_{p,q} \\ \text{(since } \sigma(p) = q)}} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)} \\
&= \sum_{p=1}^{n} \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma(p) = q}} (-1)^{\sigma} a_{p,q} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)}}_{= a_{p,q} \sum_{\sigma \in S_n; \ \sigma(p) = q} (-1)^{\sigma} \prod_{i \in \{1,2,\ldots,n\}; \ i \neq p} a_{i,\sigma(i)}} \\
&= \sum_{p=1}^{n} a_{p,q} \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma(p) = q}} (-1)^{\sigma} \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq p}} a_{i,\sigma(i)}}_{\substack{= (-1)^{p+q} \det(A_{\sim p, \sim q}) \\ \text{(by Lemma 6.84)}}} \\
&= \sum_{p=1}^{n} \underbrace{a_{p,q} (-1)^{p+q}}_{= (-1)^{p+q} a_{p,q}} \det\left(A_{\sim p, \sim q}\right) = \sum_{p=1}^{n} (-1)^{p+q} a_{p,q} \det\left(A_{\sim p, \sim q}\right).
\end{aligned}$$

This proves Theorem 6.82 **(b)**. □

Let me make three simple observations (which can easily be checked by the reader):

- Theorem 6.82 **(b)** could be (alternatively) proven using Theorem 6.82 **(a)** (applied to A^T and $a_{q,p}$ instead of A and $a_{p,q}$) and Exercise 6.4.
- Theorem 6.43 is a particular case of Theorem 6.82 **(a)**.
- Corollary 6.45 is a particular case of Theorem 6.82 **(b)**.

Remark 6.85. Some books use Laplace expansion to define the notion of a determinant. For example, one can define the determinant of a square matrix recursively, by setting the determinant of the 0×0-matrix to be 1, and defining the determinant of an $n \times n$-matrix $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (with $n > 0$) to be $\sum\limits_{q=1}^{n} (-1)^{1+q} a_{1,q} \det\left(A_{\sim 1, \sim q}\right)$ (assuming that determinants of $(n-1) \times (n-1)$-matrices such as $A_{\sim 1, \sim q}$ are already defined). Of course, this leads to the same notion of determinant as the one we are using, because of Theorem 6.82 **(a)**.

6.13. Tridiagonal determinants

In this section, we shall study the so-called *tridiagonal matrices*: a class of matrices whose all entries are zero everywhere except in the "direct proximity" of the diagonal (more specifically: on the diagonal and "one level below and one level above"). We shall find recursive formulas for the determinants of these matrices. These formulas are a simple example of an application of Laplace expansion, but also interesting in their own right.

Definition 6.86. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_{n-1}$ be $n-1$ elements of $\mathbb{K}$ (where we take the position that "-1 elements of $\mathbb{K}$" means "no elements of $\mathbb{K}$"). Let $c_1, c_2, \ldots, c_{n-1}$ be $n-1$ elements of $\mathbb{K}$. We now set

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \cdots & 0 & 0 & 0 \\ c_1 & a_2 & b_2 & \cdots & 0 & 0 & 0 \\ 0 & c_2 & a_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-2} & b_{n-2} & 0 \\ 0 & 0 & 0 & \cdots & c_{n-2} & a_{n-1} & b_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & c_{n-1} & a_n \end{pmatrix}.$$

(More formally,

$$A = \left(\begin{cases} a_i, & \text{if } i = j; \\ b_i, & \text{if } i = j-1; \\ c_j, & \text{if } i = j+1; \\ 0, & \text{otherwise} \end{cases} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

)

The matrix A is called a *tridiagonal matrix*.

We shall keep the notations n, $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_{n-1}$, $c_1, c_2, \ldots, c_{n-1}$ and A fixed for the rest of Section 6.13.

Playing around with small examples, one soon notices that the determinants of tridiagonal matrices are too complicated to have neat explicit formulas in full generality. For $n \in \{0, 1, 2, 3\}$, the determinants look as follows:

$$\det A = \det (\text{the } 0 \times 0\text{-matrix}) = 1 \qquad \text{if } n = 0;$$
$$\det A = \det \begin{pmatrix} a_1 \end{pmatrix} = a_1 \qquad \text{if } n = 1;$$
$$\det A = \det \begin{pmatrix} a_1 & b_1 \\ c_1 & a_2 \end{pmatrix} = a_1 a_2 - b_1 c_1 \qquad \text{if } n = 2;$$
$$\det A = \det \begin{pmatrix} a_1 & b_1 & 0 \\ c_1 & a_2 & b_2 \\ 0 & c_2 & a_3 \end{pmatrix} = a_1 a_2 a_3 - a_1 b_2 c_2 - a_3 b_1 c_1 \qquad \text{if } n = 3.$$

(And these formulas get more complicated the larger n becomes.) However, the many zeroes present in a tridiagonal matrix make it easy to find a recursive formula for its determinant using Laplace expansion:

Proposition 6.87. For every two elements x and y of $\{0, 1, \ldots, n\}$ satisfying $x \leq y$, we let $A_{x,y}$ be the $(y-x) \times (y-x)$-matrix

$$\begin{pmatrix} a_{x+1} & b_{x+1} & 0 & \cdots & 0 & 0 & 0 \\ c_{x+1} & a_{x+2} & b_{x+2} & \cdots & 0 & 0 & 0 \\ 0 & c_{x+2} & a_{x+3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{y-2} & b_{y-2} & 0 \\ 0 & 0 & 0 & \cdots & c_{y-2} & a_{y-1} & b_{y-1} \\ 0 & 0 & 0 & \cdots & 0 & c_{y-1} & a_y \end{pmatrix} = \operatorname{sub}_{x+1,x+2,\ldots,y}^{x+1,x+2,\ldots,y} A.$$

(a) We have $\det (A_{x,x}) = 1$ for every $x \in \{0, 1, \ldots, n\}$.

(b) We have $\det (A_{x,x+1}) = a_{x+1}$ for every $x \in \{0, 1, \ldots, n-1\}$.

(c) For every $x \in \{0, 1, \ldots, n\}$ and $y \in \{0, 1, \ldots, n\}$ satisfying $x \leq y - 2$, we have

$$\det (A_{x,y}) = a_y \det (A_{x,y-1}) - b_{y-1} c_{y-1} \det (A_{x,y-2}).$$

(d) For every $x \in \{0, 1, \ldots, n\}$ and $y \in \{0, 1, \ldots, n\}$ satisfying $x \leq y - 2$, we have

$$\det (A_{x,y}) = a_{x+1} \det (A_{x+1,y}) - b_{x+1} c_{x+1} \det (A_{x+2,y}).$$

(e) We have $A = A_{0,n}$.

Proof of Proposition 6.87. **(e)** The definition of $A_{0,n}$ yields

$$A_{0,n} = \begin{pmatrix} a_1 & b_1 & 0 & \cdots & 0 & 0 & 0 \\ c_1 & a_2 & b_2 & \cdots & 0 & 0 & 0 \\ 0 & c_2 & a_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-2} & b_{n-2} & 0 \\ 0 & 0 & 0 & \cdots & c_{n-2} & a_{n-1} & b_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & c_{n-1} & a_n \end{pmatrix} = A.$$

This proves Proposition 6.87 **(e)**.

(a) Let $x \in \{0, 1, \ldots, n\}$. Then, $A_{x,x}$ is an $(x-x) \times (x-x)$-matrix, thus a 0×0-matrix. Hence, its determinant is $\det(A_{x,x}) = 1$. This proves Proposition 6.87 **(a)**.

(b) Let $x \in \{0, 1, \ldots, n-1\}$. The definition of $A_{x,x+1}$ shows that $A_{x,x+1}$ is the 1×1-matrix $\begin{pmatrix} a_{x+1} \end{pmatrix}$. Hence, $\det(A_{x,x+1}) = \det \begin{pmatrix} a_{x+1} \end{pmatrix} = a_{x+1}$. This proves Proposition 6.87 **(b)**.

(c) Let $x \in \{0, 1, \ldots, n\}$ and $y \in \{0, 1, \ldots, n\}$ be such that $x \leq y - 2$. We have

$$A_{x,y} = \begin{pmatrix} a_{x+1} & b_{x+1} & 0 & \cdots & 0 & 0 & 0 \\ c_{x+1} & a_{x+2} & b_{x+2} & \cdots & 0 & 0 & 0 \\ 0 & c_{x+2} & a_{x+3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{y-2} & b_{y-2} & 0 \\ 0 & 0 & 0 & \cdots & c_{y-2} & a_{y-1} & b_{y-1} \\ 0 & 0 & 0 & \cdots & 0 & c_{y-1} & a_y \end{pmatrix}. \qquad (428)$$

This is a $(y-x) \times (y-x)$-matrix. If we cross out its $(y-x)$-th row (i.e., its last row) and its $(y-x)$-th column (i.e., its last column), then we obtain $A_{x,y-1}$. In other words, $(A_{x,y})_{\sim(y-x),\sim(y-x)} = A_{x,y-1}$.

Let us write the matrix $A_{x,y}$ in the form $A_{x,y} = (u_{i,j})_{1 \leq i \leq y-x,\ 1 \leq j \leq y-x}$. Thus,

$$\begin{aligned} & (u_{y-x,1}, u_{y-x,2}, \ldots, u_{y-x,y-x}) \\ & = (\text{the last row of the matrix } A_{x,y}) = (0, 0, \ldots, 0, c_{y-1}, a_y). \end{aligned}$$

In other words, we have

$$\begin{aligned} & (u_{y-x,q} = 0 \qquad \text{for every } q \in \{1, 2, \ldots, y-x-2\}), \qquad (429) \\ & u_{y-x,y-x-1} = c_{y-1}, \qquad \text{and} \\ & u_{y-x,y-x} = a_y. \end{aligned}$$

Now, Laplace expansion along the $(y-x)$-th row (or, more precisely, Theorem

6.82 **(a)**, applied to $y-x$, $A_{x,y}$, $u_{i,j}$ and $y-x$ instead of n, A, $a_{i,j}$ and p) yields

$$\begin{aligned}
\det\left(A_{x,y}\right) &= \sum_{q=1}^{y-x} (-1)^{(y-x)+q} u_{y-x,q} \det\left(\left(A_{x,y}\right)_{\sim(y-x),\sim q}\right) \\
&= \sum_{q=1}^{y-x-2} (-1)^{(y-x)+q} \underbrace{u_{y-x,q}}_{\substack{=0 \\ \text{(by (429))}}} \det\left(\left(A_{x,y}\right)_{\sim(y-x),\sim q}\right) \\
&\qquad + \underbrace{(-1)^{(y-x)+(y-x-1)}}_{=-1} \underbrace{u_{y-x,y-x-1}}_{=c_{y-1}} \det\left(\left(A_{x,y}\right)_{\sim(y-x),\sim(y-x-1)}\right) \\
&\qquad + \underbrace{(-1)^{(y-x)+(y-x)}}_{=1} \underbrace{u_{y-x,y-x}}_{=a_y} \det\left(\underbrace{\left(A_{x,y}\right)_{\sim(y-x),\sim(y-x)}}_{=A_{x,y-1}}\right) \\
&\qquad \left(\text{since } y-x \geq 2 \text{ (since } x \leq y-2)\right) \\
&= \underbrace{\sum_{q=1}^{y-x-2} (-1)^{(y-x)+q} 0 \det\left(\left(A_{x,y}\right)_{\sim(y-x),\sim q}\right)}_{=0} \\
&\qquad - c_{y-1} \det\left(\left(A_{x,y}\right)_{\sim(y-x),\sim(y-x-1)}\right) + a_y \det\left(A_{x,y-1}\right) \\
&= -c_{y-1} \det\left(\left(A_{x,y}\right)_{\sim(y-x),\sim(y-x-1)}\right) + a_y \det\left(A_{x,y-1}\right). \qquad (430)
\end{aligned}$$

Now, let $B = \left(A_{x,y}\right)_{\sim(y-x),\sim(y-x-1)}$. Thus, (430) becomes

$$\begin{aligned}
\det\left(A_{x,y}\right) &= -c_{y-1} \det\left(\underbrace{\left(A_{x,y}\right)_{\sim(y-x),\sim(y-x-1)}}_{=B}\right) + a_y \det\left(A_{x,y-1}\right) \\
&= -c_{y-1} \det B + a_y \det\left(A_{x,y-1}\right). \qquad (431)
\end{aligned}$$

Now,

$$\begin{aligned}
B &= \left(A_{x,y}\right)_{\sim(y-x),\sim(y-x-1)} \\
&= \begin{pmatrix}
a_{x+1} & b_{x+1} & 0 & \cdots & 0 & 0 & 0 \\
c_{x+1} & a_{x+2} & b_{x+2} & \cdots & 0 & 0 & 0 \\
0 & c_{x+2} & a_{x+3} & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & a_{y-3} & b_{y-3} & 0 \\
0 & 0 & 0 & \cdots & c_{y-3} & a_{y-2} & 0 \\
0 & 0 & 0 & \cdots & 0 & c_{y-2} & b_{y-1}
\end{pmatrix} \qquad \left(\text{because of (428)}\right).
\end{aligned}$$

(432)

Now, let us write the matrix B in the form $B = (v_{i,j})_{1 \leq i \leq y-x-1,\ 1 \leq j \leq y-x-1}$. Thus,

$$\begin{aligned}
&\left(v_{1,y-x-1}, v_{2,y-x-1}, \ldots, v_{y-x-1,y-x-1}\right)^T \\
&= (\text{the last column of the matrix } B) = \left(0,0,\ldots,0,b_{y-1}\right)^T
\end{aligned}$$

(because of (432)). In other words, we have

$$\left(v_{p,y-x-1} = 0 \qquad \text{for every } p \in \{1,2,\ldots,y-x-2\}\right), \qquad \text{and} \tag{433}$$
$$v_{y-x-1,y-x-1} = b_{y-1}.$$

Now, Laplace expansion along the $(y-x-1)$-th column (or, more precisely, Theorem 6.82 **(b)**, applied to $y-x-1$, B, $v_{i,j}$ and $y-x-1$ instead of n, A, $a_{i,j}$ and q) yields

$$\begin{aligned}
\det B &= \sum_{p=1}^{y-x-1} (-1)^{p+(y-x-1)} v_{p,y-x-1} \det\left(B_{\sim p,\sim(y-x-1)}\right) \\
&= \sum_{p=1}^{y-x-2} (-1)^{p+(y-x-1)} \underbrace{v_{p,y-x-1}}_{\substack{=0 \\ \text{(by (433))}}} \det\left(B_{\sim p,\sim(y-x-1)}\right) \\
&\qquad + \underbrace{(-1)^{(y-x-1)+(y-x-1)}}_{=1} \underbrace{v_{y-x-1,y-x-1}}_{=b_{y-1}} \det\left(B_{\sim(y-x-1),\sim(y-x-1)}\right) \\
&\qquad \left(\text{since } y-x-1 \geq 1 \text{ (since } x \leq y-2)\right) \\
&= \underbrace{\sum_{p=1}^{y-x-2} (-1)^{p+(y-x-1)} 0 \det\left(B_{\sim p,\sim(y-x-1)}\right)}_{=0} + b_{y-1} \det\left(B_{\sim(y-x-1),\sim(y-x-1)}\right) \\
&= b_{y-1} \det\left(B_{\sim(y-x-1),\sim(y-x-1)}\right).
\end{aligned} \tag{434}$$

Finally, a look at (432) reveals that

$$B_{\sim(y-x-1),\sim(y-x-1)} = \begin{pmatrix}
a_{x+1} & b_{x+1} & 0 & \cdots & 0 & 0 & 0 \\
c_{x+1} & a_{x+2} & b_{x+2} & \cdots & 0 & 0 & 0 \\
0 & c_{x+2} & a_{x+3} & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & a_{y-4} & b_{y-4} & 0 \\
0 & 0 & 0 & \cdots & c_{y-4} & a_{y-3} & b_{y-3} \\
0 & 0 & 0 & \cdots & 0 & c_{y-3} & a_{y-2}
\end{pmatrix} = A_{x,y-2}.$$

Hence, (434) becomes

$$\det B = b_{y-1} \det\left(\underbrace{B_{\sim(y-x-1),\sim(y-x-1)}}_{=A_{x,y-2}}\right) = b_{y-1} \det\left(A_{x,y-2}\right).$$

Therefore, (431) becomes

$$\begin{aligned}\det\left(A_{x,y}\right) &= -c_{y-1} \underbrace{\det B}_{=b_{y-1}\det\left(A_{x,y-2}\right)} + a_y \det\left(A_{x,y-1}\right) \\ &= -c_{y-1}b_{y-1}\det\left(A_{x,y-2}\right) + a_y \det\left(A_{x,y-1}\right) \\ &= a_y \det\left(A_{x,y-1}\right) - b_{y-1}c_{y-1}\det\left(A_{x,y-2}\right).\end{aligned}$$

This proves Proposition 6.87 **(c)**.

(d) The proof of Proposition 6.87 **(d)** is similar to the proof of Proposition 6.87 **(c)**. The main difference is that we now have to perform Laplace expansion along the 1-st row (instead of the $(y-x)$-th row) and then Laplace expansion along the 1-st column (instead of the $(y-x-1)$-th column). □

Proposition 6.87 gives us two fast recursive algorithms to compute $\det A$:

The first algorithm proceeds by recursively computing $\det(A_{0,m})$ for every $m \in \{0,1,\ldots,n\}$. This is done using Proposition 6.87 **(a)** (for $m=0$), Proposition 6.87 **(b)** (for $m=1$) and Proposition 6.87 **(c)** (to find $\det(A_{0,m})$ for $m \geq 2$ in terms of $\det(A_{0,m-1})$ and $\det(A_{0,m-2})$). The final value $\det(A_{0,n})$ is $\det A$ (by Proposition 6.87 **(e)**).

The second algorithm proceeds by recursively computing $\det(A_{m,n})$ for every $m \in \{0,1,\ldots,n\}$. This recursion goes backwards: We start with $m=n$ (where we use Proposition 6.87 **(a)**), then turn to $m=n-1$ (using Proposition 6.87 **(b)**), and then go further and further down (using Proposition 6.87 **(d)** to compute $\det(A_{m,n})$ in terms of $\det(A_{m+1,n})$ and $\det(A_{m+2,n})$).

So we have two different recursive algorithms leading to one and the same result. Whenever you have such a thing, you can package up the equivalence of the two algorithms as an exercise, and try to make it less easy by covering up the actual goal of the algorithms (in our case, computing $\det A$). In our case, this leads to the following exercise:

Exercise 6.25. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_{n-1}$ be $n-1$ elements of $\mathbb{K}$.

Define a sequence $(u_0, u_1, \ldots, u_n)$ of elements of $\mathbb{K}$ recursively by setting $u_0 = 1$, $u_1 = a_1$ and

$$u_i = a_i u_{i-1} - b_{i-1} u_{i-2} \qquad \text{for every } i \in \{2,3,\ldots,n\}.$$

Define a sequence $(v_0, v_1, \ldots, v_n)$ of elements of $\mathbb{K}$ recursively by setting $v_0 = 1$, $v_1 = a_n$ and

$$v_i = a_{n-i+1} v_{i-1} - b_{n-i+1} v_{i-2} \qquad \text{for every } i \in \{2,3,\ldots,n\}.$$

Prove that $u_n = v_n$.

This exercise generalizes IMO Shortlist 2013 problem A1[251].

Our recursive algorithms for computing $\det A$ also yield another observation: The determinant $\det A$ depends not on the $2(n-1)$ elements $b_1, b_2, \ldots, b_{n-1}, c_1, c_2, \ldots, c_{n-1}$ but only on the products $b_1c_1, b_2c_2, \ldots, b_{n-1}c_{n-1}$.

Exercise 6.26. Define $A_{x,y}$ as in Proposition 6.87. Prove that

$$\frac{\det A}{\det(A_{1,n})} = a_1 - \cfrac{b_1c_1}{a_2 - \cfrac{b_2c_2}{a_3 - \cfrac{b_3c_3}{a_4 - \cfrac{}{\ddots - \cfrac{b_{n-2}c_{n-2}}{a_{n-1} - \cfrac{b_{n-1}c_{n-1}}{a_n}}}}}},$$

provided that all denominators in this equality are invertible.

Exercise 6.27. Assume that $a_i = 1$ for all $i \in \{1, 2, \ldots, n\}$. Also, assume that $b_i = 1$ and $c_i = -1$ for all $i \in \{1, 2, \ldots, n-1\}$. Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence (defined as in Chapter 4). Show that $\det A = f_{n+1}$.

Remark 6.88. Consider once again the Fibonacci sequence $(f_0, f_1, f_2, \ldots)$ (defined as in Chapter 4). Let n be a positive integer. Combining the results of Exercise

[251] I have a suspicion that IMO Shortlist 2009 problem C3 also can be viewed as an equality between two recursive ways to compute a determinant; but this determinant seems to be harder to find (I don't think it can be obtained from Proposition 6.87).

6.26 and Exercise 6.27 (the details are left to the reader), we obtain the equality

$$\frac{f_{n+1}}{f_n} = 1 - \cfrac{1(-1)}{1 - \cfrac{1(-1)}{1 - \cfrac{1(-1)}{1 - \cfrac{\ddots}{-\cfrac{1(-1)}{1-\cfrac{1(-1)}{1}}}}}} \qquad \text{(with } n-1 \text{ fractions in total)}$$

$$= 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{\ddots}{+\cfrac{1}{1+\cfrac{1}{1}}}}}} \qquad \text{(with } n-1 \text{ fractions in total)}.$$

If you know some trivia about the golden ratio, you might recognize this as a part of the continued fraction for the golden ratio φ. The whole continued fraction for φ is

$$\varphi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{}{\ddots}}}} \qquad \text{(with infinitely many fractions)}.$$

This hints at the fact that $\lim\limits_{n\to\infty} \frac{f_{n+1}}{f_n} = \varphi$. (This is easy to prove without continued fractions, of course.)

6.14. On block-triangular matrices

Definition 6.89. Let n, n', m and m' be four nonnegative integers.

Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ be an $n \times m$-matrix.

Let $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m'}$ be an $n \times m'$-matrix.

Let $C = \left(c_{i,j}\right)_{1\leq i\leq n',\ 1\leq j\leq m}$ be an $n' \times m$-matrix.

Let $D = \left(d_{i,j}\right)_{1\leq i\leq n',\ 1\leq j\leq m'}$ be an $n' \times m'$-matrix.

Then, $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ will mean the $(n+n') \times (m+m')$-matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} & b_{1,1} & b_{1,2} & \cdots & b_{1,m'} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} & b_{2,1} & b_{2,2} & \cdots & b_{2,m'} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} & b_{n,1} & b_{n,2} & \cdots & b_{n,m'} \\ c_{1,1} & c_{1,2} & \cdots & c_{1,m} & d_{1,1} & d_{1,2} & \cdots & d_{1,m'} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} & d_{2,1} & d_{2,2} & \cdots & d_{2,m'} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n',1} & c_{n',2} & \cdots & c_{n',m} & d_{n',1} & d_{n',2} & \cdots & d_{n',m'} \end{pmatrix}.$$

(Formally speaking, this means that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \left(\begin{cases} a_{i,j}, & \text{if } i \leq n \text{ and } j \leq m; \\ b_{i,j-m}, & \text{if } i \leq n \text{ and } j > m; \\ c_{i-n,j}, & \text{if } i > n \text{ and } j \leq m; \\ d_{i-n,j-m}, & \text{if } i > n \text{ and } j > m \end{cases} \right)_{1 \leq i \leq n+n',\ 1 \leq j \leq m+m'} . \qquad (435)$$

Less formally, we can say that $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is the matrix obtained by gluing the matrices A, B, C and D to form one big $(n+n') \times (m+m')$-matrix, where the right border of A is glued together with the left border of B, the bottom border of A is glued together with the top border of C, etc.)

Do not get fooled by the notation $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$: It is (in general) not a 2×2-matrix, but an $(n+n') \times (m+m')$-matrix, and its entries are not A, B, C and D but the entries of A, B, C and D.

Example 6.90. If $A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$, $B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, $C = \begin{pmatrix} c_1 & c_2 \end{pmatrix}$ and $D = \begin{pmatrix} d \end{pmatrix}$, then $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & b_1 \\ a_{2,1} & a_{2,2} & b_2 \\ c_1 & c_2 & d \end{pmatrix}$.

The notation $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ introduced in Definition 6.89 is a particular case of a more general notation – the *block-matrix construction* – for gluing together multiple matrices with matching dimensions[252]. We shall only need the particular case that is Definition 6.89, however.

Definition 6.91. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Recall that $\mathbb{K}^{n\times m}$ is the set of all $n \times m$-matrices.

We use $0_{n\times m}$ (or sometimes just 0) to denote the $n \times m$ *zero matrix*. (As we recall, this is the $n \times m$-matrix whose all entries are 0; in other words, this is the $n \times m$-matrix $(0)_{1\leq i\leq n,\ 1\leq j\leq m}$.)

Exercise 6.28. Let n, n', m, m', ℓ and ℓ' be six nonnegative integers. Let $A \in \mathbb{K}^{n\times m}$, $B \in \mathbb{K}^{n\times m'}$, $C \in \mathbb{K}^{n'\times m}$, $D \in \mathbb{K}^{n'\times m'}$, $A' \in \mathbb{K}^{m\times \ell}$, $B' \in \mathbb{K}^{m\times \ell'}$, $C' \in \mathbb{K}^{m'\times \ell}$ and $D' \in \mathbb{K}^{m'\times \ell'}$. Then, prove that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} AA' + BC' & AB' + BD' \\ CA' + DC' & CB' + DD' \end{pmatrix}.$$

Remark 6.92. The intuitive meaning of Exercise 6.28 is that the product of two matrices in "block-matrix notation" can be computed by applying the usual multiplication rule "on the level of blocks", without having to fall back to multiplying single entries. However, when applying Exercise 6.28, do not forget to check that its conditions are satisfied. Let me give an example and a non-example:

Example: If $A = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, $B = \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{pmatrix}$, $C = \begin{pmatrix} c \end{pmatrix}$, $D = \begin{pmatrix} d_1 & d_2 \end{pmatrix}$, $A' = \begin{pmatrix} a'_1 & a'_2 \end{pmatrix}$, $B' = \begin{pmatrix} b'_1 & b'_2 \end{pmatrix}$, $C' = \begin{pmatrix} c'_{1,1} & c'_{1,2} \\ c'_{2,1} & c'_{2,2} \end{pmatrix}$ and $D' = \begin{pmatrix} d'_{1,1} & d'_{1,2} \\ d'_{2,1} & d'_{2,2} \end{pmatrix}$,

[252] This construction defines an $(n_1 + n_2 + \cdots + n_x) \times (m_1 + m_2 + \cdots + m_y)$-matrix

$$\begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,y} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,y} \\ \vdots & \vdots & \ddots & \vdots \\ A_{x,1} & A_{x,2} & \cdots & A_{x,y} \end{pmatrix} \tag{436}$$

whenever you have given two nonnegative integers x and y, an x-tuple $(n_1, n_2, \ldots, n_x) \in \mathbb{N}^x$, a y-tuple $(m_1, m_2, \ldots, m_y) \in \mathbb{N}^y$, and an $n_i \times m_j$-matrix $A_{i,j}$ for every $i \in \{1, 2, \ldots, x\}$ and every $j \in \{1, 2, \ldots, y\}$. I guess you can guess the definition of this matrix. So you start with an "$x \times y$-matrix of matrices" and glue them together to an $(n_1 + n_2 + \cdots + n_x) \times (m_1 + m_2 + \cdots + m_y)$-matrix (provided that the dimensions of these matrices allow them to be glued – e.g., you cannot glue a 2×3-matrix to a 4×6-matrix along its right border, nor on any other border).

It is called "block-matrix construction" because the original matrices $A_{i,j}$ appear as "blocks" in the big matrix (436). Most authors define block matrices to be matrices which are "partitioned" into blocks as in (436); this is essentially our construction in reverse: Instead of gluing several "small" matrices into a big one, they study big matrices partitioned into many small matrices. Of course, the properties of their "block matrices" are equivalent to those of our "block-matrix construction".

then Exercise 6.28 can be applied (with $n = 3$, $n' = 1$, $m = 1$, $m' = 2$, $\ell = 2$ and $\ell' = 2$), and thus we obtain

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} AA' + BC' & AB' + BD' \\ CA' + DC' & CB' + DD' \end{pmatrix}.$$

Non-example: If $A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $B = \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix}$, $C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, $D = \begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \end{pmatrix}$, $A' = \begin{pmatrix} a'_{1,1} & a'_{1,2} \\ a'_{2,1} & a'_{2,2} \end{pmatrix}$, $B' = \begin{pmatrix} b'_{1,1} & b'_{1,2} \\ b'_{2,1} & b'_{2,2} \end{pmatrix}$, $C' = \begin{pmatrix} c'_1 & c'_2 \end{pmatrix}$ and $D' = \begin{pmatrix} d'_1 & d'_2 \end{pmatrix}$, then Exercise 6.28 cannot be applied, because there exist no $n, m, \ell \in \mathbb{N}$ such that $A \in \mathbb{K}^{n \times m}$ and $A' \in \mathbb{K}^{m \times \ell}$. (Indeed, the number of columns of A does not equal the number of rows of A', but these numbers would both have to be m.) The matrices $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ and $\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$ still exist in this case, and can even be multiplied, but their product is not given by a simple formula such as the one in Exercise 6.28. Thus, beware of seeing Exercise 6.28 as a panacea for multiplying matrices blockwise.

Exercise 6.29. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let B be an $n \times m$-matrix. Let D be an $m \times m$-matrix. Prove that

$$\det \begin{pmatrix} A & B \\ 0_{m \times n} & D \end{pmatrix} = \det A \cdot \det D.$$

Example 6.93. Exercise 6.29 (applied to $n = 2$ and $m = 3$) yields

$$\det \begin{pmatrix} a_{1,1} & a_{1,2} & b_{1,1} & b_{1,2} & b_{1,3} \\ a_{2,1} & a_{2,2} & b_{2,1} & b_{2,2} & b_{2,3} \\ 0 & 0 & c_{1,1} & c_{1,2} & c_{1,3} \\ 0 & 0 & c_{2,1} & c_{2,2} & c_{2,3} \\ 0 & 0 & c_{3,1} & c_{3,2} & c_{3,3} \end{pmatrix} = \det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \cdot \det \begin{pmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,1} & c_{3,2} & c_{3,3} \end{pmatrix}.$$

Remark 6.94. Not every determinant of the form $\det \begin{pmatrix} A & B \\ 0_{m \times n} & D \end{pmatrix}$ can be computed using Exercise 6.29. In fact, Exercise 6.29 requires A to be an $n \times n$-matrix and D to be an $m \times m$-matrix; thus, both A and D have to be square matrices in order for Exercise 6.29 to be applicable. For instance, Exercise 6.29 cannot be applied to compute $\det \begin{pmatrix} a_1 & b_{1,1} & b_{1,2} \\ a_2 & b_{2,1} & b_{2,2} \\ 0 & c_1 & c_2 \end{pmatrix}$.

Remark 6.95. You might wonder whether Exercise 6.29 generalizes to a formula for $\det\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ when $A \in \mathbb{K}^{n\times n}$, $B \in \mathbb{K}^{n\times m}$, $C \in \mathbb{K}^{m\times n}$ and $D \in \mathbb{K}^{m\times m}$. The general answer is "No". However, when D is invertible, there exists such a formula (the Schur complement formula shown in Exercise 6.36 below). Curiously, there is also a formula for the case when $n = m$ and $CD = DC$ (see [Silves00, Theorem 3]).

We notice that Exercise 6.29 allows us to solve Exercise 6.6 in a new way.

An analogue of Exercise 6.29 exists in which the $0_{m\times n}$ in the lower-left part of the matrix is replaced by a $0_{n\times m}$ in the upper-right part:

Exercise 6.30. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let C be an $m \times n$-matrix. Let D be an $m \times m$-matrix. Prove that

$$\det\begin{pmatrix} A & 0_{n\times m} \\ C & D \end{pmatrix} = \det A \cdot \det D.$$

Exercise 6.31. (a) Compute the determinant of the 7×7-matrix

$$\begin{pmatrix} a & 0 & 0 & 0 & 0 & 0 & b \\ 0 & a' & 0 & 0 & 0 & b' & 0 \\ 0 & 0 & a'' & 0 & b'' & 0 & 0 \\ 0 & 0 & 0 & e & 0 & 0 & 0 \\ 0 & 0 & c'' & 0 & d'' & 0 & 0 \\ 0 & c' & 0 & 0 & 0 & d' & 0 \\ c & 0 & 0 & 0 & 0 & 0 & d \end{pmatrix},$$

where $a, a', a'', b, b', b'', c, c', c'', d, d', d'', e$ are elements of $\mathbb{K}$.

(b) Compute the determinant of the 6×6-matrix

$$\begin{pmatrix} a & 0 & 0 & \ell & 0 & 0 \\ 0 & b & 0 & 0 & m & 0 \\ 0 & 0 & c & 0 & 0 & n \\ g & 0 & 0 & d & 0 & 0 \\ 0 & h & 0 & 0 & e & 0 \\ 0 & 0 & k & 0 & 0 & f \end{pmatrix},$$

where $a, b, c, d, e, f, g, h, k, \ell, m, n$ are elements of $\mathbb{K}$.

Exercise 6.32. Invent and solve an exercise on computing determinants.

6.15. The adjugate matrix

We start this section with a variation on Theorem 6.82:

Proposition 6.96. Let $n \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n \times n$-matrix. Let $r \in \{1, 2, \ldots, n\}$.

(a) For every $p \in \{1, 2, \ldots, n\}$ satisfying $p \neq r$, we have

$$0 = \sum_{q=1}^{n} (-1)^{p+q} a_{r,q} \det\left(A_{\sim p,\sim q}\right).$$

(b) For every $q \in \{1, 2, \ldots, n\}$ satisfying $q \neq r$, we have

$$0 = \sum_{p=1}^{n} (-1)^{p+q} a_{p,r} \det\left(A_{\sim p,\sim q}\right).$$

Proof of Proposition 6.96. **(a)** Let $p \in \{1, 2, \ldots, n\}$ be such that $p \neq r$.

Let C be the $n \times n$-matrix obtained from A by replacing the p-th row of A by the r-th row of A. Thus, the p-th and the r-th rows of C are equal. Therefore, the matrix C has two equal rows (since $p \neq r$). Hence, $\det C = 0$ (by Exercise 6.7 **(e)**, applied to C instead of A).

Let us write the $n \times n$-matrix C in the form $C = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

The p-th row of C equals the r-th row of A (by the construction of C). In other words,

$$c_{p,q} = a_{r,q} \qquad \text{for every } q \in \{1, 2, \ldots, n\}. \tag{437}$$

On the other hand, the matrix C equals the matrix A in all rows but the p-th one (again, by the construction of C). Hence, if we cross out the p-th rows in both C and A, then the matrices C and A become equal. Therefore,

$$C_{\sim p,\sim q} = A_{\sim p,\sim q} \qquad \text{for every } q \in \{1, 2, \ldots, n\} \tag{438}$$

(because the construction of $C_{\sim p,\sim q}$ from C involves crossing out the p-th row, and so does the construction of $A_{\sim p,\sim q}$ from A).

Now, $\det C = 0$, so that

$$\begin{aligned} 0 = \det C &= \sum_{q=1}^{n} (-1)^{p+q} \underbrace{c_{p,q}}_{\substack{=a_{r,q} \\ \text{(by (437))}}} \det\left(\underbrace{C_{\sim p,\sim q}}_{\substack{=A_{\sim p,\sim q} \\ \text{(by (438))}}}\right) \\ &\qquad \left(\text{by Theorem 6.82 **(a)**, applied to } C \text{ and } c_{i,j} \text{ instead of } A \text{ and } a_{i,j}\right) \\ &= \sum_{q=1}^{n} (-1)^{p+q} a_{r,q} \det\left(A_{\sim p,\sim q}\right). \end{aligned}$$

This proves Proposition 6.96 **(a)**.

(b) This proof is rather similar to the proof of Proposition 6.96 **(a)**, except that rows are now replaced by columns. We leave the details to the reader. □

We now can define the "adjugate" of a matrix:

Definition 6.97. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. We define a new $n \times n$-matrix $\operatorname{adj} A$ by

$$\operatorname{adj} A = \left((-1)^{i+j} \det \left(A_{\sim j, \sim i} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

This matrix $\operatorname{adj} A$ is called the *adjugate* of the matrix A. (Some authors call it the "adjunct" or "adjoint" or "classical adjoint" of A instead. However, beware of the word "adjoint": It means too many different things; in particular it has a second meaning for a matrix.)

The appearance of $A_{\sim j, \sim i}$ (not $A_{\sim i, \sim j}$) in Definition 6.97 might be surprising, but it is not a mistake. We will soon see what it is good for.

There is also a related notion, namely that of a "cofactor matrix". The *cofactor matrix* of an $n \times n$-matrix A is defined to be $\left((-1)^{i+j} \det \left(A_{\sim i, \sim j} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. This is, of course, the transpose $(\operatorname{adj} A)^T$ of $\operatorname{adj} A$. The entries of this matrix are called the *cofactors* of A.

Example 6.98. The adjugate of the 0×0-matrix is the 0×0-matrix.

The adjugate of a 1×1-matrix $\begin{pmatrix} a \end{pmatrix}$ is $\operatorname{adj} \begin{pmatrix} a \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix}$. (Yes, this shows that all 1×1-matrices have the same adjugate.)

The adjugate of a 2×2-matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is

$$\operatorname{adj} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The adjugate of a 3×3-matrix $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ is

$$\operatorname{adj} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - ge & bg - ah & ae - bd \end{pmatrix}.$$

Proposition 6.99. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then, $\operatorname{adj} \left(A^T \right) = (\operatorname{adj} A)^T$.

Proof of Proposition 6.99. Let $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$.
From $i \in \{1, 2, \ldots, n\}$, we obtain $1 \leq i \leq n$, so that $n \geq 1$ and thus $n - 1 \in \mathbb{N}$.

The definition of $A_{\sim i,\sim j}$ yields $A_{\sim i,\sim j} = \operatorname{sub}^{1,2,\ldots,\widehat{j},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,n} A$. But the definition of $\left(A^T\right)_{\sim j,\sim i}$ yields

$$\left(A^T\right)_{\sim j,\sim i} = \operatorname{sub}^{1,2,\ldots,\widehat{i},\ldots,n}_{1,2,\ldots,\widehat{j},\ldots,n} \left(A^T\right). \tag{439}$$

On the other hand, Proposition 6.79 **(e)** (applied to $m = n$, $u = n-1$, $v = n-1$, $(i_1, i_2, \ldots, i_u) = \left(1, 2, \ldots, \widehat{i}, \ldots, n\right)$ and $(j_1, j_2, \ldots, j_v) = \left(1, 2, \ldots, \widehat{j}, \ldots, n\right)$) yields $\left(\operatorname{sub}^{1,2,\ldots,\widehat{j},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,n} A\right)^T = \operatorname{sub}^{1,2,\ldots,\widehat{i},\ldots,n}_{1,2,\ldots,\widehat{j},\ldots,n} \left(A^T\right)$. Compared with (439), this yields

$$\left(A^T\right)_{\sim j,\sim i} = \left(\underbrace{\operatorname{sub}^{1,2,\ldots,\widehat{j},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,n} A}_{=A_{\sim i,\sim j}}\right)^T = \left(A_{\sim i,\sim j}\right)^T.$$

Hence,

$$\det\left(\underbrace{\left(A^T\right)_{\sim j,\sim i}}_{=\left(A_{\sim i,\sim j}\right)^T}\right) = \det\left(\left(A_{\sim i,\sim j}\right)^T\right) = \det\left(A_{\sim i,\sim j}\right) \tag{440}$$

(by Exercise 6.4, applied to $n-1$ and $A_{\sim i,\sim j}$ instead of n and A).

Let us now forget that we fixed i and j. We thus have shown that (440) holds for every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$.

Now, $\operatorname{adj} A = \left((-1)^{i+j} \det\left(A_{\sim j,\sim i}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, and thus the definition of the transpose of a matrix shows that

$$(\operatorname{adj} A)^T = \left(\underbrace{(-1)^{j+i}}_{=(-1)^{i+j}} \det\left(A_{\sim i,\sim j}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n} = \left((-1)^{i+j} \det\left(A_{\sim i,\sim j}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Compared with

$$\begin{aligned}
\operatorname{adj}\left(A^T\right) &= \left((-1)^{i+j} \underbrace{\det\left(\left(A^T\right)_{\sim j,\sim i}\right)}_{\substack{=\det\left(A_{\sim i,\sim j}\right) \\ \text{(by (440))}}}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \\
&\qquad \left(\text{by the definition of } \operatorname{adj}\left(A^T\right)\right) \\
&= \left((-1)^{i+j} \det\left(A_{\sim i,\sim j}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n},
\end{aligned}$$

this yields $\operatorname{adj}\left(A^T\right) = (\operatorname{adj} A)^T$. This proves Proposition 6.99. □

The most important property of adjugates, however, is the following fact:

Theorem 6.100. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then,

$$A \cdot \operatorname{adj} A = \operatorname{adj} A \cdot A = \det A \cdot I_n.$$

(Recall that I_n denotes the $n \times n$ identity matrix. Expressions such as $\operatorname{adj} A \cdot A$ and $\det A \cdot I_n$ have to be understood as $(\operatorname{adj} A) \cdot A$ and $(\det A) \cdot I_n$, respectively.)

Example 6.101. Recall that the adjugate of a 2×2-matrix is given by the formula $\operatorname{adj}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Thus, Theorem 6.100 (applied to $n = 2$) yields

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot I_2.$$

(Of course, $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot I_2 = (ad - bc) \cdot I_2 = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}$.)

Proof of Theorem 6.100. For any two objects i and j, we define $\delta_{i,j}$ to be the element $\begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$ of $\mathbb{K}$. Then, $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of I_n), and thus

$$\det A \cdot \underbrace{I_n}_{= \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}} = \det A \cdot \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(\det A \cdot \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}. \tag{441}$$

On the other hand, let us write the matrix A in the form $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Then, the definition of the product of two matrices shows that

$$\begin{aligned}
& A \cdot \operatorname{adj} A \\
&= \left(\sum_{k=1}^{n} a_{i,k} (-1)^{k+j} \det \left(A_{\sim j, \sim k} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \\
&\qquad \left(\begin{array}{c} \text{since } A = \left(a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \\ \text{and } \operatorname{adj} A = \left((-1)^{i+j} \det \left(A_{\sim j, \sim i} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \end{array} \right) \\
&= \left(\sum_{q=1}^{n} \underbrace{a_{i,q} (-1)^{q+j}}_{= (-1)^{q+j} a_{i,q}} \det \left(A_{\sim j, \sim q} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \\
&\qquad \left(\text{here, we renamed the summation index } k \text{ as } q \right) \\
&= \left(\sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det \left(A_{\sim j, \sim q} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} . \qquad (442)
\end{aligned}$$

Now, we claim that

$$\sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det \left(A_{\sim j, \sim q} \right) = \det A \cdot \delta_{i,j} \qquad (443)$$

for any $(i, j) \in \{1, 2, \ldots, n\}^2$.

[*Proof of (443):* Fix $(i, j) \in \{1, 2, \ldots, n\}^2$. We are in one of the following two cases:

Case 1: We have $i = j$.

Case 2: We have $i \neq j$.

Let us consider Case 1 first. In this case, we have $i = j$. Hence, $\delta_{i,j} = 1$. Now, Theorem 6.82 **(a)** (applied to $p = i$) yields

$$\det A = \sum_{q=1}^{n} \underbrace{(-1)^{i+q}}_{\substack{= (-1)^{q+i} = (-1)^{q+j} \\ (\text{since } i = j)}} a_{i,q} \det \left(\underbrace{A_{\sim i, \sim q}}_{\substack{= A_{\sim j, \sim q} \\ (\text{since } i = j)}} \right) = \sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det \left(A_{\sim j, \sim q} \right) .$$

In view of $\det A \cdot \underbrace{\delta_{i,j}}_{=1} = \det A$, this rewrites as

$$\det A \cdot \delta_{i,j} = \sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det \left(A_{\sim j, \sim q} \right) .$$

Thus, (443) is proven in Case 1.

Let us next consider Case 2. In this case, we have $i \neq j$. Hence, $\delta_{i,j} = 0$ and $j \neq i$. Now, Proposition 6.96 **(a)** (applied to $p = j$ and $r = i$) yields

$$0 = \sum_{q=1}^{n} \underbrace{(-1)^{j+q}}_{=(-1)^{q+j}} a_{i,q} \det\left(A_{\sim j,\sim q}\right) = \sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det\left(A_{\sim j,\sim q}\right).$$

In view of $\det A \cdot \underbrace{\delta_{i,j}}_{=0} = 0$, this rewrites as

$$\det A \cdot \delta_{i,j} = \sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det\left(A_{\sim j,\sim q}\right).$$

Thus, (443) is proven in Case 2.

We have now proven (443) in each of the two Cases 1 and 2. Thus, (443) is proven.]

Now, (442) becomes

$$A \cdot \operatorname{adj} A = \left(\underbrace{\sum_{q=1}^{n} (-1)^{q+j} a_{i,q} \det\left(A_{\sim j,\sim q}\right)}_{\substack{=\det A \cdot \delta_{i,j} \\ \text{(by (443))}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$$
$$= \left(\det A \cdot \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \det A \cdot I_n \qquad (444)$$

(by (441)).

It now remains to prove that $\operatorname{adj} A \cdot A = \det A \cdot I_n$. One way to do this is by mimicking the above proof using Theorem 6.82 **(b)** and Proposition 6.96 **(b)** instead of Theorem 6.82 **(a)** and Proposition 6.96 **(a)**. However, here is a slicker proof:

Let us forget that we fixed A. We thus have shown that (444) holds for every $n \times n$-matrix A.

Now, let A be any $n \times n$-matrix. Then, we can apply (444) to A^T instead of A. We thus obtain

$$A^T \cdot \operatorname{adj}\left(A^T\right) = \underbrace{\det\left(A^T\right)}_{\substack{=\det A \\ \text{(by Exercise 6.4)}}} \cdot I_n = \det A \cdot I_n. \qquad (445)$$

Now, (348) (applied to $u = n$, $v = n$, $w = n$, $P = \operatorname{adj} A$ and $Q = A$) shows that

$$(\operatorname{adj} A \cdot A)^T = A^T \cdot \underbrace{(\operatorname{adj} A)^T}_{\substack{=\operatorname{adj}\left(A^T\right) \\ \text{(by Proposition 6.99)}}} = A^T \cdot \operatorname{adj}\left(A^T\right) = \det A \cdot I_n \qquad \text{(by (445))}.$$

Hence,

$$\left(\underbrace{(\operatorname{adj} A \cdot A)^T}_{=\det A \cdot I_n}\right)^T = (\det A \cdot I_n)^T = \det A \cdot \underbrace{(I_n)^T}_{\substack{=I_n \\ \text{(by (349), applied to } u=n)}}$$

$$\left(\text{by (350), applied to } u = n,\ v = n,\ P = I_n \text{ and } \lambda = \det A\right)$$

$$= \det A \cdot I_n.$$

Compared with

$$\left((\operatorname{adj} A \cdot A)^T\right)^T = \operatorname{adj} A \cdot A \qquad \left(\text{by (351), applied to } u = n,\ v = n \text{ and } P = \operatorname{adj} A \cdot A\right),$$

this yields $\operatorname{adj} A \cdot A = \det A \cdot I_n$. Combined with (444), this yields

$$A \cdot \operatorname{adj} A = \operatorname{adj} A \cdot A = \det A \cdot I_n.$$

This proves Theorem 6.100. □

The following is a simple consequence of Theorem 6.100:

Corollary 6.102. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let v be a column vector with n entries. If $Av = 0_{n\times 1}$, then $\det A \cdot v = 0_{n\times 1}$.

(Recall that $0_{n\times 1}$ denotes the $n \times 1$ zero matrix, i.e., the column vector with n entries whose all entries are 0.)

Proof of Corollary 6.102. Assume that $Av = 0_{n\times 1}$. It is easy to see that every $m \in \mathbb{N}$ and every $n \times m$-matrix B satisfy $I_n B = B$. Applying this to $m = 1$ and $B = v$, we obtain $I_n v = v$.

It is also easy to see that every $m \in \mathbb{N}$ and every $m \times n$-matrix B satisfy $B \cdot 0_{n\times 1} = 0_{m\times 1}$. Applying this to $m = n$ and $B = \operatorname{adj} A$, we obtain $\operatorname{adj} A \cdot 0_{n\times 1} = 0_{n\times 1}$.

Now, Theorem 6.100 yields $\operatorname{adj} A \cdot A = \det A \cdot I_n$. Hence,

$$\underbrace{(\operatorname{adj} A \cdot A)}_{=\det A \cdot I_n} v = (\det A \cdot I_n)\, v = \det A \cdot \underbrace{(I_n v)}_{=v} = \det A \cdot v.$$

Compared to

$$(\operatorname{adj} A \cdot A)\, v = \operatorname{adj} A \cdot \underbrace{(Av)}_{=0_{n\times 1}} \qquad \text{(since matrix multiplication is associative)}$$

$$= \operatorname{adj} A \cdot 0_{n\times 1} = 0_{n\times 1},$$

this yields $\det A \cdot v = 0_{n\times 1}$. This proves Corollary 6.102. □

Exercise 6.33. Let $n \in \mathbb{N}$. Let A and B be two $n \times n$-matrices. Prove that

$$\operatorname{adj}(AB) = \operatorname{adj} B \cdot \operatorname{adj} A.$$

Let me end this section with another application of Proposition 6.96:

Exercise 6.34. Let $n \in \mathbb{N}$. For every n elements $y_1, y_2, \ldots, y_n$ of $\mathbb{K}$, we define an element $V(y_1, y_2, \ldots, y_n)$ of $\mathbb{K}$ by

$$V(y_1, y_2, \ldots, y_n) = \prod_{1 \leq i < j \leq n} (y_i - y_j).$$

Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. Prove that

$$\sum_{k=1}^{n} x_k V(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n) = \left(\binom{n}{2} t + \sum_{k=1}^{n} x_k \right) V(x_1, x_2, \ldots, x_n).$$

[**Hint:** Use Theorem 6.46, Laplace expansion, Proposition 6.96 and the binomial formula.]

Exercise 6.34 is part of [Fulton97, §4.3, Exercise 10].

6.16. Inverting matrices

We now will study inverses of matrices. We begin with a definition:

Definition 6.103. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix.

(a) A *left inverse* of A means an $m \times n$-matrix L such that $LA = I_m$. We say that the matrix A is *left-invertible* if and only if a left inverse of A exists.

(b) A *right inverse* of A means an $m \times n$-matrix R such that $AR = I_n$. We say that the matrix A is *right-invertible* if and only if a right inverse of A exists.

(c) An *inverse* of A (or *two-sided inverse* of A) means an $m \times n$-matrix B such that $BA = I_m$ and $AB = I_n$. We say that the matrix A is *invertible* if and only if an inverse of A exists.

The notions "left-invertible", "right-invertible" and "invertible" depend on the ring $\mathbb{K}$. We shall therefore speak of "left-invertible over $\mathbb{K}$", "right-invertible over $\mathbb{K}$" and "invertible over $\mathbb{K}$" whenever the context does not unambiguously determine $\mathbb{K}$.

The notions of "left inverse", "right inverse" and "inverse" are not interchangeable (unlike for elements in a commutative ring). We shall soon see in what cases they are identical; but first, let us give a few examples.

Example 6.104. For this example, set $\mathbb{K} = \mathbb{Z}$.

Let P be the 1×2-matrix $\begin{pmatrix} 1 & 2 \end{pmatrix}$. The matrix P is right-invertible. For instance, $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$ are two right inverses of P (because $P \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} = I_1$ and $P \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} = I_1$). This example shows that the right inverse of a matrix is not always unique.

The 2×1-matrix $P^T = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is left-invertible. The left inverses of P^T are the transposes of the right inverses of P.

The matrix P is not left-invertible; the matrix P^T is not right-invertible.

Let Q be the 2×2-matrix $\begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix}$. The matrix Q is invertible. Its inverse is $\begin{pmatrix} -2 & 1 \\ -3 & 1 \end{pmatrix}$ (since $\begin{pmatrix} -2 & 1 \\ -3 & 1 \end{pmatrix} Q = I_2$ and $Q \begin{pmatrix} -2 & 1 \\ -3 & 1 \end{pmatrix} = I_2$). It is not hard to see that this is its only inverse.

Let R be the 2×2-matrix $\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$. It can be seen that this matrix is not invertible **as a matrix over** $\mathbb{Z}$. On the other hand, if we consider it as a matrix over $\mathbb{K} = \mathbb{Q}$ instead, then it is invertible, with inverse $\begin{pmatrix} 1/5 & 2/5 \\ 2/5 & -1/5 \end{pmatrix}$.

Of course, any inverse of a matrix A is automatically both a left inverse of A and a right inverse of A. Thus, an invertible matrix A is automatically both left-invertible and right-invertible.

The following simple fact is an analogue of Proposition 6.65:

Proposition 6.105. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let L be a left inverse of A. Let R be a right inverse of A.

(a) We have $L = R$.

(b) The matrix A is invertible, and $L = R$ is an inverse of A.

Proof of Proposition 6.105. We know that L is a left inverse of A. In other words, L is an $m \times n$-matrix such that $LA = I_m$ (by the definition of a "left inverse").

We know that R is a right inverse of A. In other words, R is an $m \times n$-matrix such that $AR = I_n$ (by the definition of a "right inverse").

Now, recall that $I_m G = G$ for every $k \in \mathbb{N}$ and every $m \times k$-matrix G. Applying this to $k = n$ and $G = R$, we obtain $I_m R = R$.

Also, recall that $G I_n = G$ for every $k \in \mathbb{N}$ and every $k \times n$-matrix G. Applying this to $k = m$ and $G = L$, we obtain $L I_n = L$. Thus, $L = L \underbrace{I_n}_{=AR} = \underbrace{LA}_{=I_m} R = I_m R = R$. This proves Proposition 6.105 **(a)**.

(b) We have $LA = I_m$ and $A \underbrace{L}_{=R} = AR = I_n$. Thus, L is an $m \times n$-matrix such that $LA = I_m$ and $AL = I_n$. In other words, L is an inverse of A (by the definition of an "inverse"). Thus, $L = R$ is an inverse of A (since $L = R$). This proves Proposition 6.105 **(b)**. □

Corollary 6.106. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix.
(a) If A is left-invertible and right-invertible, then A is invertible.
(b) If A is invertible, then there exists exactly one inverse of A.

Proof of Corollary 6.106. **(a)** Assume that A is left-invertible and right-invertible. Thus, A has a left inverse L (since A is left-invertible). Consider this L. Also, A has a right inverse R (since A is right-invertible). Consider this R. Proposition 6.105 **(b)** yields that the matrix A is invertible, and $L = R$ is an inverse of A. Corollary 6.106 **(a)** is proven.

(b) Assume that A is invertible. Let B and B' be any two inverses of A. Since B is an inverse of A, we know that B is an $m \times n$-matrix such that $BA = I_m$ and $AB = I_n$ (by the definition of an "inverse"). Thus, in particular, B is an $m \times n$-matrix such that $BA = I_m$. In other words, B is a left inverse of A. Since B' is an inverse of A, we know that B' is an $m \times n$-matrix such that $B'A = I_m$ and $AB' = I_n$ (by the definition of an "inverse"). Thus, in particular, B' is an $m \times n$-matrix such that $AB' = I_n$. In other words, B' is a right inverse of A. Now, Proposition 6.105 **(a)** (applied to $L = B$ and $R = B'$) shows that $B = B'$.

Let us now forget that we fixed B and B'. We thus have shown that if B and B' are two inverses of A, then $B = B'$. In other words, any two inverses of A are equal. In other words, there exists at most one inverse of A. Since we also know that there exists at least one inverse of A (since A is invertible), we thus conclude that there exists exactly one inverse of A. This proves Corollary 6.106 **(b)**. □

Definition 6.107. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an invertible $n \times m$-matrix. Corollary 6.106 **(b)** shows that there exists exactly one inverse of A. Thus, we can speak of "*the inverse of* A". We denote this inverse by A^{-1}.

In contrast to Definition 6.66, we do **not** define the notation B/A for two matrices B and A for which A is invertible. In fact, the trouble with such a notation would be its ambiguity: should it mean BA^{-1} or $A^{-1}B$? (In general, BA^{-1} and $A^{-1}B$ are not the same.) Some authors do write B/A for the matrices BA^{-1} and $A^{-1}B$ when these matrices are equal; but we shall not have a reason to do so.

Remark 6.108. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an invertible $n \times m$-matrix. Then, the inverse A^{-1} of A is an $m \times n$-matrix and satisfies $AA^{-1} = I_n$ and $A^{-1}A = I_m$. This follows from the definition of the inverse of A; we are just stating it once again, because it will later be used without mention.

Example 6.104 (and your experiences with a linear algebra class, if you have taken one) suggest the conjecture that only square matrices can be invertible. Indeed, this is **almost** true. There is a stupid counterexample: If $\mathbb{K}$ is a trivial ring, then every matrix over $\mathbb{K}$ is invertible[253]. It turns out that this is the only case where nonsquare matrices can be invertible. Indeed, we have the following:

Theorem 6.109. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix.

(a) If A is left-invertible and if $n < m$, then $\mathbb{K}$ is a trivial ring.

(b) If A is right-invertible and if $n > m$, then $\mathbb{K}$ is a trivial ring.

(c) If A is invertible and if $n \neq m$, then $\mathbb{K}$ is a trivial ring.

Proof of Theorem 6.109. **(a)** Assume that A is left-invertible, and that $n < m$.

The matrix A has a left inverse L (since it is left-invertible). Consider this L.

We know that L is a left inverse of A. In other words, L is an $m \times n$-matrix such that $LA = I_m$ (by the definition of a "left inverse"). But (379) (applied to m, n, L and A instead of n, m, A and B) yields $\det(LA) = 0$ (since $n < m$). Thus, $0 = \det\left(\underbrace{LA}_{=I_m}\right) = \det(I_m) = 1$. Of course, the 0 and the 1 in this equality mean the elements $0_{\mathbb{K}}$ and $1_{\mathbb{K}}$ of $\mathbb{K}$ (rather than the integers 0 and 1); thus, it rewrites as $0_{\mathbb{K}} = 1_{\mathbb{K}}$. In other words, $\mathbb{K}$ is a trivial ring. This proves Theorem 6.109 **(a)**.

(b) Assume that A is right-invertible, and that $n > m$.

The matrix A has a right inverse R (since it is right-invertible). Consider this R.

We know that R is a right inverse of A. In other words, R is an $m \times n$-matrix such that $AR = I_n$ (by the definition of a "right inverse"). But (379) (applied to $B = R$) yields $\det(AR) = 0$ (since $m < n$). Thus, $0 = \det\left(\underbrace{AR}_{=I_n}\right) = \det(I_n) = 1$. Of course, the 0 and the 1 in this equality mean the elements $0_{\mathbb{K}}$ and $1_{\mathbb{K}}$ of $\mathbb{K}$ (rather than the integers 0 and 1); thus, it rewrites as $0_{\mathbb{K}} = 1_{\mathbb{K}}$. In other words, $\mathbb{K}$ is a trivial ring. This proves Theorem 6.109 **(b)**.

(c) Assume that A is invertible, and that $n \neq m$. Since $n \neq m$, we must be in one of the following two cases:

Case 1: We have $n < m$.

Case 2: We have $n > m$.

Let us first consider Case 1. In this case, we have $n < m$. Now, A is invertible, and thus left-invertible (since every invertible matrix is left-invertible). Hence, $\mathbb{K}$ is a trivial ring (according to Theorem 6.109 **(a)**). Thus, Theorem 6.109 **(c)** is proven in Case 1.

[253]For example, the 1×2-matrix $\begin{pmatrix} 0_{\mathbb{K}} & 0_{\mathbb{K}} \end{pmatrix}$ over a trivial ring $\mathbb{K}$ is invertible, having inverse $\begin{pmatrix} 0_{\mathbb{K}} \\ 0_{\mathbb{K}} \end{pmatrix}$. If you don't believe me, just check that

$$\begin{pmatrix} 0_{\mathbb{K}} \\ 0_{\mathbb{K}} \end{pmatrix} \begin{pmatrix} 0_{\mathbb{K}} & 0_{\mathbb{K}} \end{pmatrix} = \begin{pmatrix} 0_{\mathbb{K}} & 0_{\mathbb{K}} \\ 0_{\mathbb{K}} & 0_{\mathbb{K}} \end{pmatrix} = \begin{pmatrix} 1_{\mathbb{K}} & 0_{\mathbb{K}} \\ 0_{\mathbb{K}} & 1_{\mathbb{K}} \end{pmatrix} \qquad (\text{since } 0_{\mathbb{K}} = 1_{\mathbb{K}})$$
$$= I_2$$

and $\begin{pmatrix} 0_{\mathbb{K}} & 0_{\mathbb{K}} \end{pmatrix} \begin{pmatrix} 0_{\mathbb{K}} \\ 0_{\mathbb{K}} \end{pmatrix} = \begin{pmatrix} 0_{\mathbb{K}} \end{pmatrix} = \begin{pmatrix} 1_{\mathbb{K}} \end{pmatrix} = I_1$.

Let us now consider Case 2. In this case, we have $n > m$. Now, A is invertible, and thus right-invertible (since every invertible matrix is right-invertible). Hence, $\mathbb{K}$ is a trivial ring (according to Theorem 6.109 **(b)**). Thus, Theorem 6.109 **(c)** is proven in Case 2.

We have thus proven Theorem 6.109 **(c)** in both Cases 1 and 2. Thus, Theorem 6.109 **(c)** always holds. $\square$

Theorem 6.109 **(c)** says that the question whether a matrix is invertible is only interesting for square matrices, unless the ring $\mathbb{K}$ is given so inexplicitly that we do not know whether it is trivial or not[254]. Let us now study the invertibility of a square matrix. Here, the determinant turns out to be highly useful:

Theorem 6.110. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix.

(a) The matrix A is invertible if and only if the element $\det A$ of $\mathbb{K}$ is invertible (in $\mathbb{K}$).

(b) If $\det A$ is invertible, then the inverse of A is $A^{-1} = \dfrac{1}{\det A} \cdot \operatorname{adj} A$.

When $\mathbb{K}$ is a field, the invertible elements of $\mathbb{K}$ are precisely the nonzero elements of $\mathbb{K}$. Thus, when $\mathbb{K}$ is a field, the statement of Theorem 6.110 **(a)** can be rewritten as "The matrix A is invertible if and only if $\det A \neq 0$"; this is a cornerstone of linear algebra. But our statement of Theorem 6.110 **(a)** works for an arbitrary commutative ring $\mathbb{K}$. In particular, it works for $\mathbb{K} = \mathbb{Z}$. Here is a consequence:

Corollary 6.111. Let $n \in \mathbb{N}$. Let $A \in \mathbb{Z}^{n \times n}$ be an $n \times n$-matrix over $\mathbb{Z}$. Then, the matrix A is invertible if and only if $\det A \in \{1, -1\}$.

Proof of Corollary 6.111. If g is an integer, then g is invertible (in $\mathbb{Z}$) if and only if $g \in \{1, -1\}$. In other words, for every integer g, we have the following equivalence:

$$(g \text{ is invertible (in } \mathbb{Z})) \iff (g \in \{1, -1\}). \tag{446}$$

Now, Theorem 6.110 **(a)** (applied to $\mathbb{K} = \mathbb{Z}$) yields that the matrix A is invertible if and only if the element $\det A$ of $\mathbb{Z}$ is invertible (in $\mathbb{Z}$). Thus, we have the following chain of equivalences:

$$\begin{aligned} & (\text{the matrix } A \text{ is invertible}) \\ \iff & (\det A \text{ is invertible (in } \mathbb{Z})) \iff (\det A \in \{1, -1\}) \\ & \qquad (\text{by (446), applied to } g = \det A). \end{aligned}$$

This proves Corollary 6.111. $\square$

[254] This actually happens rather often in algebra! For example, rings are often defined by "generators and relations" (such as "the ring with commuting generators a, b, c subject to the relations $a^2 + b^2 = c^2$ and $ab = c$"). Sometimes the relations force the ring to become trivial (for instance, the ring with generator a and relations $a = 1$ and $a^2 = 2$ is clearly the trivial ring, because in this ring we have $2 = a^2 = 1^2 = 1$). Often this is not clear a-priori, and theorems such as Theorem 6.109 can be used to show this. The triviality of a ring can be a nontrivial statement! (Richman makes this point in [Richma88].)

Notice that Theorem 6.110 **(b)** yields an explicit way to compute the inverse of a square matrix A (provided that we can compute determinants and the inverse of $\det A$). This is not the fastest way (at least not when $\mathbb{K}$ is a field), but it is useful for various theoretical purposes.

Proof of Theorem 6.110. **(a)** $\Longrightarrow$: [255] Assume that the matrix A is invertible. In other words, an inverse B of A exists. Consider such a B.

The matrix B is an inverse of A. In other words, B is an $n \times n$-matrix such that $BA = I_n$ and $AB = I_n$ (by the definition of an "inverse"). Theorem 6.23 yields $\det(AB) = \det A \cdot \det B$, so that $\det A \cdot \det B = \det\left(\underbrace{AB}_{=I_n}\right) = \det(I_n) = 1$. Of course, we also have $\det B \cdot \det A = \det A \cdot \det B = 1$. Thus, $\det B$ is an inverse of $\det A$ in $\mathbb{K}$. Therefore, the element $\det A$ is invertible (in $\mathbb{K}$). This proves the $\Longrightarrow$ direction of Theorem 6.110 **(a)**.

$\Longleftarrow$: Assume that the element $\det A$ is invertible (in $\mathbb{K}$). Thus, its inverse $\dfrac{1}{\det A}$ exists. Theorem 6.100 yields

$$A \cdot \operatorname{adj} A = \operatorname{adj} A \cdot A = \det A \cdot I_n.$$

Now, define an $n \times n$-matrix B by $B = \dfrac{1}{\det A} \cdot \operatorname{adj} A$. Then,

$$A \underbrace{B}_{=\frac{1}{\det A} \cdot \operatorname{adj} A} = A \cdot \left(\frac{1}{\det A} \cdot \operatorname{adj} A\right) = \frac{1}{\det A} \cdot \underbrace{A \cdot \operatorname{adj} A}_{=\det A \cdot I_n} = \underbrace{\frac{1}{\det A} \cdot \det A}_{=1} \cdot I_n = I_n$$

and

$$\underbrace{B}_{=\frac{1}{\det A} \cdot \operatorname{adj} A} A = \frac{1}{\det A} \cdot \underbrace{\operatorname{adj} A \cdot A}_{=\det A \cdot I_n} = \underbrace{\frac{1}{\det A} \cdot \det A}_{=1} \cdot I_n = I_n.$$

Thus, B is an $n \times n$-matrix such that $BA = I_n$ and $AB = I_n$. In other words, B is an inverse of A (by the definition of an "inverse"). Thus, an inverse of A exists; in other words, the matrix A is invertible. This proves the $\Longleftarrow$ direction of Theorem 6.110 **(a)**.

[255] In case you don't know what the notation "$\Longrightarrow$:" here means:

Theorem 6.110 **(a)** is an "if and only if" assertion. In other words, it asserts that $\mathcal{U} \Longleftrightarrow \mathcal{V}$ for two statements $\mathcal{U}$ and $\mathcal{V}$. (In our case, $\mathcal{U}$ is the statement "the matrix A is invertible", and $\mathcal{V}$ is the statement "the element $\det A$ of $\mathbb{K}$ is invertible (in $\mathbb{K}$)".) In order to prove a statement of the form $\mathcal{U} \Longleftrightarrow \mathcal{V}$, it is sufficient to prove the implications $\mathcal{U} \Longrightarrow \mathcal{V}$ and $\mathcal{U} \Longleftarrow \mathcal{V}$. Usually, these two implications are proven separately (although not always; for instance, in the proof of Corollary 6.111, we have used a chain of equivalences to prove $\mathcal{U} \Longleftrightarrow \mathcal{V}$ directly). When writing such a proof, one often uses the abbreviations "$\Longrightarrow$:" and "$\Longleftarrow$:" for "Here comes the proof of the implication $\mathcal{U} \Longrightarrow \mathcal{V}$:" and "Here comes the proof of the implication $\mathcal{U} \Longleftarrow \mathcal{V}$:", respectively.

We have now proven both directions of Theorem 6.110 **(a)**. Theorem 6.110 **(a)** is thus proven.

(b) Assume that $\det A$ is invertible. Thus, its inverse $\dfrac{1}{\det A}$ exists. We define an $n \times n$-matrix B by $B = \dfrac{1}{\det A} \cdot \operatorname{adj} A$. Then, B is an inverse of A [256]. In other words, B is **the** inverse of A. In other words, $B = A^{-1}$. Hence, $A^{-1} = B = \dfrac{1}{\det A} \cdot \operatorname{adj} A$. This proves Theorem 6.110 **(b)**. □

Corollary 6.112. Let $n \in \mathbb{N}$. Let A and B be two $n \times n$-matrices such that $AB = I_n$.

(a) We have $BA = I_n$.

(b) The matrix A is invertible, and the matrix B is the inverse of A.

Proof of Corollary 6.112. Theorem 6.23 yields $\det(AB) = \det A \cdot \det B$, so that $\det A \cdot \det B = \det\left(\underbrace{AB}_{=I_n}\right) = \det(I_n) = 1$. Of course, we also have $\det B \cdot \det A = \det A \cdot \det B = 1$. Thus, $\det B$ is an inverse of $\det A$ in $\mathbb{K}$. Therefore, the element $\det A$ is invertible (in $\mathbb{K}$). Therefore, the matrix A is invertible (according to the $\Longleftarrow$ direction of Theorem 6.110 **(b)**). Thus, the inverse of A exists. Let C be this inverse. Thus, C is a left inverse of A (since every inverse of A is a left inverse of A).

The matrix B is an $n \times n$-matrix satisfying $AB = I_n$. In other words, B is a right inverse of A. On the other hand, C is a left inverse of A. Hence, Proposition 6.105 **(a)** (applied to $L = C$ and $R = B$) yields $C = B$. Hence, the matrix B is the inverse of A (since the matrix C is the inverse of A). Thus, Corollary 6.112 **(b)** is proven.

Since B is the inverse of A, we have $BA = I_n$ and $AB = I_n$ (by the definition of an "inverse"). This proves Corollary 6.112 **(a)**. □

Remark 6.113. Corollary 6.112 is **not** obvious! Matrix multiplication, in general, is not commutative (we have $AB \neq BA$ more often than not), and there is no reason to expect that $AB = I_n$ implies $BA = I_n$. The fact that this is nevertheless true for square matrices took us quite some work to prove (we needed, among other things, the notion of an adjugate). This fact would **not** hold for rectangular matrices. Nor does it hold for "infinite square matrices": Without wanting to go into the details of how products of infinite matrices are defined, I invite you to check that the two infinite matrices $A = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$ and $B = A^T =$

[256] We have shown this in our proof of the $\Longleftarrow$ direction of Theorem 6.110 **(a)**.

$\begin{pmatrix} 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$ satisfy $AB = I_\infty$ but $BA \neq I_\infty$. This makes Corollary 6.112 **(a)** all the more interesting.

Here are some more exercises involving matrices in "block-matrix form":

Exercise 6.35. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times n}$, $B \in \mathbb{K}^{n\times m}$, $C \in \mathbb{K}^{m\times n}$ and $D \in \mathbb{K}^{m\times m}$. Furthermore, let $W \in \mathbb{K}^{m\times m}$ and $V \in \mathbb{K}^{m\times n}$ be such that $VA = -WC$. Prove that

$$\det W \cdot \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \cdot \det (VB + WD).$$

[**Hint:** Use Exercise 6.28 to simplify the product $\begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix}$; then, take determinants.]

Exercise 6.35 can often be used to compute the determinant of a matrix given in block-matrix form (i.e., determinants of the form $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix}$) by only computing determinants of smaller matrices (such as W, A and $VB + WD$). It falls short of providing a general method for computing such determinants[257], but it is one of the most general facts about them. The next two exercises are two special cases of Exercise 6.35:

Exercise 6.36. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times n}$, $B \in \mathbb{K}^{n\times m}$, $C \in \mathbb{K}^{m\times n}$ and $D \in \mathbb{K}^{m\times m}$ be such that the matrix A is invertible. Prove that

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \cdot \det \left(D - CA^{-1}B\right).$$

Exercise 6.36 is known as the *Schur complement formula* (or, at least, it is one of several formulas sharing this name); and the matrix $D - CA^{-1}B$ appearing on its right hand side is known as the *Schur complement* of the block A in the matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$.

[257]Indeed, it only gives a formula for $\det W \cdot \det \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, not for $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. If $\det W$ is invertible, then it allows for computing $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix}$; but Exercise 6.35 gives no hint on how to find matrices W and V such that $\det W$ is invertible and such that $VA = -WC$. (Actually, such matrices do not always exist!)

Exercise 6.37. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times n}$, $B \in \mathbb{K}^{n\times m}$, $C \in \mathbb{K}^{m\times n}$ and $D \in \mathbb{K}^{m\times m}$. Let $A' \in \mathbb{K}^{n\times n}$, $B' \in \mathbb{K}^{n\times m}$, $C' \in \mathbb{K}^{m\times n}$ and $D' \in \mathbb{K}^{m\times m}$. Assume that the matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is invertible, and that its inverse is the matrix $\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$. Prove that

$$\det A = \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \det \left(D'\right).$$

Exercise 6.37 can be rewritten in the following more handy form:

Exercise 6.38. We shall use the notations introduced in Definition 6.78.

Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times n}$ be an invertible matrix. Let $k \in \{0, 1, \ldots, n\}$. Prove that

$$\det \left(\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k} A\right) = \det A \cdot \det \left(\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n} \left(A^{-1}\right)\right).$$

Exercise 6.38 is a particular case of the so-called *Jacobi complementary minor theorem* (Exercise 6.56 further below).

6.17. Noncommutative rings

I think that here is a good place to introduce two other basic notions from algebra: that of a noncommutative ring, and that of a group.

Definition 6.114. The notion of a *noncommutative ring* is defined in the same way as we have defined a commutative ring (in Definition 6.2), except that we no longer require the "Commutativity of multiplication" axiom.

As I have already said, the word "noncommutative" (in "noncommutative ring") does not mean that commutativity of multiplication has to be false in this ring; it only means that commutativity of multiplication is not required. Thus, every commutative ring is a noncommutative ring. Therefore, each of the examples of a commutative ring given in Section 6.1 is also an example of a noncommutative ring. Of course, it is more interesting to see some examples of noncommutative rings which actually fail to obey commutativity of multiplication. Here are some of these examples:

- If $n \in \mathbb{N}$ and if $\mathbb{K}$ is a commutative ring, then the set $\mathbb{K}^{n\times n}$ of matrices becomes a noncommutative ring (when endowed with the addition and multiplication of matrices, with the zero $0_{n\times n}$ and with the unity I_n). This is actually a commutative ring when $\mathbb{K}$ is trivial or when $n \leq 1$, but in all "interesting" cases it is not commutative.

- If you have heard of the quaternions, you should realize that they form a noncommutative ring.

- Given a commutative ring $\mathbb{K}$ and n distinct symbols $X_1, X_2, \ldots, X_n$, we can define a *ring of polynomials in the* ***noncommutative*** *variables* $X_1, X_2, \ldots, X_n$ over $\mathbb{K}$. We do not want to go into the details of its definition at this point, but let us just mention some examples of its elements: For instance, the ring of polynomials in the noncommutative variables X and Y over $\mathbb{Q}$ contains elements such as $1 + \dfrac{2}{3}X$, $X^2 + \dfrac{3}{2}Y - 7XY + YX$, $2XY$, $2YX$ and $5X^2Y - 6XYX + 7Y^2X$ (and of course, the elements XY and YX are not equal).

- If $n \in \mathbb{N}$ and if $\mathbb{K}$ is a commutative ring, then the set of all lower-triangular $n \times n$-matrices over $\mathbb{K}$ becomes a noncommutative ring (with addition, multiplication, zero and unity defined in the same way as in $\mathbb{K}^{n \times n}$). This is because the sum and the product of any two lower-triangular $n \times n$-matrices over $\mathbb{K}$ are again lower-triangular[258], and because the matrices $0_{n \times n}$ and I_n are lower-triangular.

- In contrast, the set of all invertible 2×2-matrices over $\mathbb{K}$ is **not** a noncommutative ring (for example, because the sum of the two invertible matrices I_2 and $-I_2$ is not invertible[259]).

- If $\mathbb{K}$ is a commutative ring, then the set of all 3×3-matrices (over $\mathbb{K}$) of the form $\begin{pmatrix} a & b & c \\ 0 & d & 0 \\ 0 & 0 & f \end{pmatrix}$ (with $a, b, c, d, f \in \mathbb{K}$) is a noncommutative ring (again, with the same addition, multiplication, zero and unity as for $\mathbb{K}^{n \times n}$). [260]

- On the other hand, if $\mathbb{K}$ is a commutative ring, then the set of all 3×3-matrices (over $\mathbb{K}$) of the form $\begin{pmatrix} a & b & 0 \\ 0 & c & d \\ 0 & 0 & f \end{pmatrix}$ (with $a, b, c, d, f \in \mathbb{K}$) is **not** a noncommutative ring (unless $\mathbb{K}$ is trivial), because products of matrices in

[258]Check this! (For the sum, it is clear, but for the product, it is an instructive exercise.)

[259]unless the ring $\mathbb{K}$ is trivial

[260]To check this, one needs to prove that the matrices $0_{3 \times 3}$ and I_3 have this form, and that the sum and the product of any two matrices of this form is again a matrix of this form. All of this is clear, except for the claim about the product. The latter claim follows from the computation

$$\begin{pmatrix} a & b & c \\ 0 & d & 0 \\ 0 & 0 & f \end{pmatrix} \begin{pmatrix} a' & b' & c' \\ 0 & d' & 0 \\ 0 & 0 & f' \end{pmatrix} = \begin{pmatrix} aa' & bd' + ab' & cf' + ac' \\ 0 & dd' & 0 \\ 0 & 0 & ff' \end{pmatrix}.$$

this set are not always in this set[261].

For the rest of this section, we let $\mathbb{L}$ be a **noncommutative** ring. What can we do with elements of $\mathbb{L}$? We can do some of the things that we can do with a commutative ring, but not all of them. For example, we can still define the sum $a_1 + a_2 + \cdots + a_n$ and the product $a_1 a_2 \cdots a_n$ of n elements of a noncommutative ring. But we cannot arbitrarily reorder the factors of a product and expect to always get the same result! (With a sum, we can do this.) We can still define na for any $n \in \mathbb{Z}$ and $a \in \mathbb{L}$ (in the same way as we defined na for $n \in \mathbb{Z}$ and $a \in \mathbb{K}$ when $\mathbb{K}$ was a commutative ring). We can still define a^n for any $n \in \mathbb{N}$ and $a \in \mathbb{L}$ (again, in the same fashion as for commutative rings). The identities (327), (330), (331), (332), (333), (334), (335) and (336) still hold when the commutative ring $\mathbb{K}$ is replaced by the noncommutative ring $\mathbb{L}$; but the identities (337) and (338) may not (although they **do** hold if we additionally assume that $ab = ba$). Finite sums such as $\sum\limits_{s \in S} a_s$ (where S is a finite set, and $a_s \in \mathbb{L}$ for every $s \in S$) are well-defined, but finite products such as $\prod\limits_{s \in S} a_s$ are not (unless we specify the order in which their factors are to be multiplied).

We can define matrices over $\mathbb{L}$ in the same way as we have defined matrices over $\mathbb{K}$. We can even define the determinant of a square matrix over $\mathbb{L}$ using the formula (340); however, this determinant lacks many of the important properties that determinants over $\mathbb{K}$ have (for instance, it satisfies neither Exercise 6.4 nor Theorem 6.23), and is therefore usually not studied.[262]

We define the notion of an *inverse* of an element $a \in \mathbb{L}$; in order to do so, we simply replace $\mathbb{K}$ by $\mathbb{L}$ in Definition 6.64. (Now it suddenly matters that we required both $ab = 1$ and $ba = 1$ in Definition 6.64.) Proposition 6.65 still holds (and its proof still works) when $\mathbb{K}$ is replaced by $\mathbb{L}$.

We define the notion of an *invertible element* of $\mathbb{L}$; in order to do so, we simply replace $\mathbb{K}$ by $\mathbb{L}$ in Definition 6.66 **(a)**. We cannot directly replace $\mathbb{K}$ by $\mathbb{L}$ in Definition 6.66 **(b)**, because for two invertible elements a and b of $\mathbb{L}$ we do not necessarily have $(ab)^{-1} = a^{-1} b^{-1}$; but something very similar holds (namely, $(ab)^{-1} = b^{-1} a^{-1}$). Trying to generalize Definition 6.66 **(c)** to noncommutative rings is rather hopeless: In general, we cannot bring a "noncommutative fraction" of the form $ba^{-1} + dc^{-1}$ to a "common denominator".

Example 6.115. Let $\mathbb{K}$ be a commutative ring. Let $n \in \mathbb{N}$. As we know, $\mathbb{K}^{n \times n}$ is a noncommutative ring. The invertible elements of this ring are exactly the invertible $n \times n$-matrices. (To see this, just compare the definition of an invert-

[261] Indeed, $\begin{pmatrix} a & b & 0 \\ 0 & c & d \\ 0 & 0 & f \end{pmatrix} \begin{pmatrix} a' & b' & 0 \\ 0 & c' & d' \\ 0 & 0 & f' \end{pmatrix} = \begin{pmatrix} aa' & ab' + bc' & bd' \\ 0 & cc' & cd' + df' \\ 0 & 0 & ff' \end{pmatrix}$ can have $bd' \neq 0$.

[262] Some algebraists have come up with subtler notions of determinants for matrices over noncommutative rings. But I don't want to go in that direction here.

ible element of $\mathbb{K}^{n\times n}$ with the definition of an invertible $n \times n$-matrix. These definitions are clearly equivalent.)

6.18. Groups, and the group of units

Let me finally define the notion of a *group*.

Definition 6.116. A *group* means a set G endowed with

- a binary operation called "multiplication" (or "composition", or just "binary operation"), and denoted by $\cdot$, and written infix, and
- an element called 1_G (or e_G)

such that the following axioms are satisfied:

- *Associativity:* We have $a(bc) = (ab)c$ for all $a \in G$, $b \in G$ and $c \in G$. Here and in the following, the expression "ab" is shorthand for "$a \cdot b$" (as is usual for products of numbers).
- *Neutrality of* 1*:* We have $a1_G = 1_G a = a$ for all $a \in G$.
- *Existence of inverses:* For every $a \in G$, there exists an element $a' \in G$ such that $aa' = a'a = 1_G$. This a' is commonly denoted by a^{-1} and called the *inverse* of a. (It is easy to check that it is unique.)

Definition 6.117. The element 1_G of a group G is denoted the *neutral element* (or the *identity*) of G.

The binary operation $\cdot$ in Definition 6.116 is usually not identical with the binary operation $\cdot$ on the set of integers, and is denoted by $\cdot_G$ when confusion can arise.

The definition of a group has similarities with that of a noncommutative ring. Viewed from a distance, it may look as if a noncommutative ring would "consist" of two groups with the same underlying set. This is not quite correct, though, because the multiplication in a nontrivial ring does not satisfy the "existence of inverses" axiom. But it is true that there are two groups in every noncommutative ring:

Proposition 6.118. Let $\mathbb{L}$ be a noncommutative ring.

(a) The set $\mathbb{L}$, endowed with the **addition** $+_{\mathbb{L}}$ (as multiplication) and the element $0_{\mathbb{L}}$ (as neutral element), is a group. This group is called the *additive group* of $\mathbb{L}$, and denoted by $\mathbb{L}^+$.

(b) Let $\mathbb{L}^\times$ denote the set of all invertible elements of $\mathbb{L}$. Then, the product of two elements of $\mathbb{L}^\times$ again belongs to $\mathbb{L}^\times$. Thus, we can define a binary operation $\cdot_{\mathbb{L}^\times}$ on the set $\mathbb{L}^\times$ (written infix) by

$$a \cdot_{\mathbb{L}^\times} b = ab \qquad \text{for all } a \in \mathbb{L}^\times \text{ and } b \in \mathbb{L}^\times.$$

The set $\mathbb{L}^\times$, endowed with the multiplication $\cdot_{\mathbb{L}^\times}$ (as multiplication) and the element $1_\mathbb{L}$ (as neutral element), is a group. This group is called the *group of units* of $\mathbb{L}$.

Proof of Proposition 6.118. **(a)** The addition $+_\mathbb{L}$ is clearly a binary operation on $\mathbb{L}$, and the element $0_\mathbb{L}$ is clearly an element of $\mathbb{L}$. The three axioms in Definition 6.116 are clearly satisfied for the binary operation $+_\mathbb{L}$ and the element $0_\mathbb{L}$ [263]. Therefore, the set $\mathbb{L}$, endowed with the addition $+_\mathbb{L}$ (as multiplication) and the element $0_\mathbb{L}$ (as neutral element), is a group. This proves Proposition 6.118 **(a)**.

(b) If $a \in \mathbb{L}^\times$ and $b \in \mathbb{L}^\times$, then $ab \in \mathbb{L}^\times$ [264]. In other words, the product of two elements of $\mathbb{L}^\times$ again belongs to $\mathbb{L}^\times$. Thus, we can define a binary operation $\cdot_{\mathbb{L}^\times}$ on the set $\mathbb{L}^\times$ (written infix) by

$$a \cdot_{\mathbb{L}^\times} b = ab \qquad \text{for all } a \in \mathbb{L}^\times \text{ and } b \in \mathbb{L}^\times.$$

Also, $1_\mathbb{L}$ is an invertible element of $\mathbb{L}$ (indeed, its inverse is $1_\mathbb{L}$), and thus an element of $\mathbb{L}^\times$.

Now, we need to prove that the set $\mathbb{L}^\times$, endowed with the multiplication $\cdot_{\mathbb{L}^\times}$ (as multiplication) and the element $1_\mathbb{L}$ (as neutral element), is a group. In order to do so, we need to check that the "associativity", "neutrality of 1" and "existence of inverses" axioms are satisfied.

The "associativity" axiom follows from the "associativity of multiplication" axiom in the definition of a noncommutative ring. The "neutrality of 1" axiom follows from the "unitality" axiom in the definition of a noncommutative ring. It thus remains to prove that the "existence of inverses" axiom holds.

[263] In fact, they boil down to the "associativity of addition", "neutrality of 0" and "existence of additive inverses" axioms in the definition of a noncommutative ring.

[264] *Proof.* Let $a \in \mathbb{L}^\times$ and $b \in \mathbb{L}^\times$. We have $a \in \mathbb{L}^\times$; in other words, a is an invertible element of $\mathbb{L}$ (because $\mathbb{L}^\times$ is the set of all invertible elements of $\mathbb{L}$). Thus, the inverse a^{-1} of a is well-defined. Similarly, the inverse b^{-1} of b is well-defined. Now, since we have

$$\left(b^{-1}a^{-1}\right)(ab) = b^{-1} \underbrace{a^{-1}a}_{=1_\mathbb{L}} b = b^{-1}b = 1_\mathbb{L}$$

and

$$(ab)\left(b^{-1}a^{-1}\right) = a \underbrace{bb^{-1}}_{=1_\mathbb{L}} a^{-1} = aa^{-1} = 1_\mathbb{L},$$

we see that the element $b^{-1}a^{-1}$ of $\mathbb{L}$ is an inverse of ab. Thus, the element ab has an inverse. In other words, ab is invertible. In other words, $ab \in \mathbb{L}^\times$ (since $\mathbb{L}^\times$ is the set of all invertible elements of $\mathbb{L}$), qed.

Thus, let $a \in \mathbb{L}^\times$. We need to show that there exists an $a' \in \mathbb{L}^\times$ such that $a \cdot_{\mathbb{L}^\times} a' = a' \cdot_{\mathbb{L}^\times} a = 1_{\mathbb{L}}$ (since $1_{\mathbb{L}}$ is the neutral element of $\mathbb{L}^\times$).

We know that a is an invertible element of $\mathbb{L}$ (since $a \in \mathbb{L}^\times$); it thus has an inverse a^{-1}. Now, a itself is an inverse of a^{-1} (since $aa^{-1} = 1_{\mathbb{L}}$ and $a^{-1}a = 1_{\mathbb{L}}$), and thus the element a^{-1} of $\mathbb{L}$ has an inverse. In other words, a^{-1} is invertible, so that $a^{-1} \in \mathbb{L}^\times$. The definition of the operation $\cdot_{\mathbb{L}^\times}$ shows that $a \cdot_{\mathbb{L}^\times} a^{-1} = aa^{-1} = 1_{\mathbb{L}}$ and that $a^{-1} \cdot_{\mathbb{L}^\times} a = a^{-1}a = 1_{\mathbb{L}}$. Hence, there exists an $a' \in \mathbb{L}^\times$ such that $a \cdot_{\mathbb{L}^\times} a' = a' \cdot_{\mathbb{L}^\times} a = 1_{\mathbb{L}}$ (namely, $a' = a^{-1}$). Thus we have proven that the "existence of inverses" axiom holds. The proof of Proposition 6.118 **(b)** is thus complete. □

We now have a plentitude of examples of groups: For every noncommutative ring $\mathbb{L}$, we have the two groups $\mathbb{L}^+$ and $\mathbb{L}^\times$ defined in Proposition 6.118. Another example, for every set X, is the symmetric group of X (endowed with the composition of permutations as multiplication, and the identity permutation $\mathrm{id} : X \to X$ as the neutral element). (Many other examples can be found in textbooks on algebra, such as [Artin10] or [Goodma15]. Groups also naturally appear in the analysis of puzzles such as Rubik's cube; this is explained in various sources such as [Mulhol16], [Bump02] and [Joyner08], which can also be read as introductions to groups.)

Remark 6.119. Throwing all notational ballast aside, we can restate Proposition 6.118 **(b)** as follows: The set of all invertible elements of a noncommutative ring $\mathbb{L}$ is a group (where the binary operation is multiplication). We can apply this to the case where $\mathbb{L} = \mathbb{K}^{n\times n}$ for a commutative ring $\mathbb{K}$ and an integer $n \in \mathbb{N}$. Thus, we obtain that the set of all invertible elements of $\mathbb{K}^{n\times n}$ is a group. Since we know that the invertible elements of $\mathbb{K}^{n\times n}$ are exactly the invertible $n \times n$-matrices (by Example 6.115), we thus have shown that the set of all invertible $n \times n$-matrices is a group. This group is commonly denoted by $\mathrm{GL}_n(\mathbb{K})$; it is called the *general linear group of degree n* over $\mathbb{K}$.

6.19. Cramer's rule

Let us return to the classical properties of determinants. We have already proven many, but here is one more: It is an application of determinants to solving systems of linear equations.

Theorem 6.120. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let $b = (b_1, b_2, \ldots, b_n)^T$ be a column vector with n entries (that is, an $n \times 1$-matrix).[265]

For every $j \in \{1, 2, \ldots, n\}$, let $A_j^\#$ be the $n \times n$-matrix obtained from A by replacing the j-th column of A with the vector b.

(a) We have $A \cdot \left(\det\left(A_1^\#\right), \det\left(A_2^\#\right), \ldots, \det\left(A_n^\#\right)\right)^T = \det A \cdot b$.

(b) Assume that the matrix A is invertible. Then,

$$A^{-1}b = \left(\frac{\det\left(A_1^\#\right)}{\det A}, \frac{\det\left(A_2^\#\right)}{\det A}, \ldots, \frac{\det\left(A_n^\#\right)}{\det A}\right)^T.$$

Theorem 6.120 (or either part of it) is known as *Cramer's rule*.

Remark 6.121. A system of n linear equations in n variables $x_1, x_2, \ldots, x_n$ can be written in the form $Ax = b$, where A is a fixed $n \times n$-matrix and b is a column vector with n entries (and where x is the column vector $(x_1, x_2, \ldots, x_n)^T$ containing all the variables). When the matrix A is invertible, it thus has a unique solution: namely, $x = A^{-1}b$ (just multiply the equation $Ax = b$ from the left with A^{-1} to see this), and this solution can be computed using Theorem 6.120. This looks nice, but isn't actually all that useful for solving systems of linear equations: For one thing, this does not immediately help us solve systems of fewer or more than n equations in n variables; and even in the case of exactly n equations, the matrix A coming from a system of linear equations will not always be invertible (and in the more interesting cases, it will not be). For another thing, at least when $\mathbb{K}$ is a field, there are faster ways to solve a system of linear equations than anything that involves computing $n+1$ determinants of $n \times n$-matrices. Theorem 6.120 nevertheless turns out to be useful in proofs.

Proof of Theorem 6.120. **(a)** Fix $j \in \{1, 2, \ldots, n\}$. Let $C = A_j^{\#}$. Thus, $C = A_j^{\#}$ is the $n \times n$-matrix obtained from A by replacing the j-th column of A with the vector b. In particular, the j-th column of C is the vector b. In other words, we have

$$c_{p,j} = b_p \qquad \text{for every } p \in \{1, 2, \ldots, n\}. \tag{447}$$

Furthermore, the matrix C is equal to the matrix A in all columns but its j-th column (because it is obtained from A by replacing the j-th column of A with the vector b). Thus, if we cross out the j-th columns in both matrices C and A, then these two matrices become equal. Consequently,

$$C_{\sim p, \sim j} = A_{\sim p, \sim j} \qquad \text{for every } p \in \{1, 2, \ldots, n\} \tag{448}$$

(because the matrices $C_{\sim p, \sim j}$ and $A_{\sim p, \sim j}$ are obtained by crossing out the p-th row

[265] The reader should keep in mind that $(b_1, b_2, \ldots, b_n)^T$ is just a space-saving way to write $\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$.

and the j-th column in the matrices C and A, respectively). Now,

$$\det\left(\underbrace{A_j^{\#}}_{=C}\right) = \det C = \sum_{p=1}^{n} (-1)^{p+j} \underbrace{c_{p,j}}_{\substack{=b_p \\ \text{(by (447))}}} \det\left(\underbrace{C_{\sim p,\sim j}}_{\substack{=A_{\sim p,\sim j} \\ \text{(by (448))}}}\right)$$

$$\left(\begin{array}{c}\text{by Theorem 6.82 } \textbf{(b)}\text{, applied} \\ \text{to } C,\ c_{i,j} \text{ and } j \text{ instead of } A,\ a_{i,j} \text{ and } q\end{array}\right)$$

$$= \sum_{p=1}^{n} (-1)^{p+j} b_p \det\left(A_{\sim p,\sim j}\right). \qquad (449)$$

Let us now forget that we fixed j. We thus have proven (449) for every $j \in \{1,2,\ldots,n\}$. Now, fix $i \in \{1,2,\ldots,n\}$. Then, for every $p \in \{1,2,\ldots,n\}$ satisfying $p \neq i$, we have

$$\sum_{q=1}^{n} b_p (-1)^{p+q} a_{i,q} \det\left(A_{\sim p,\sim q}\right) = 0 \qquad (450)$$

[266]. Also, we have

$$\sum_{q=1}^{n} b_i (-1)^{i+q} a_{i,q} \det\left(A_{\sim i,\sim q}\right) = \det A \cdot b_i \qquad (452)$$

[266] *Proof of (450):* Let $p \in \{1,2,\ldots,n\}$ be such that $p \neq i$. Hence, Proposition 6.96 **(a)** (applied to $r = i$) shows that

$$0 = \sum_{q=1}^{n} (-1)^{p+q} a_{i,q} \det\left(A_{\sim p,\sim q}\right). \qquad (451)$$

Now,

$$\sum_{q=1}^{n} b_p (-1)^{p+q} a_{i,q} \det\left(A_{\sim p,\sim q}\right) = b_p \underbrace{\sum_{q=1}^{n} (-1)^{p+q} a_{i,q} \det\left(A_{\sim p,\sim q}\right)}_{\substack{=0 \\ \text{(by (451))}}} = 0.$$

Thus, (450) is proven.

[267]. Now,

$$\sum_{k=1}^{n} a_{i,k} \underbrace{\det\left(A_k^{\#}\right)}_{\substack{=\sum_{p=1}^{n}(-1)^{p+k}b_p\det\left(A_{\sim p,\sim k}\right) \\ \text{(by (449), applied to } j=k\text{)}}}$$
$$= \sum_{k=1}^{n} a_{i,k} \sum_{p=1}^{n} (-1)^{p+k} b_p \det\left(A_{\sim p,\sim k}\right) = \sum_{q=1}^{n} a_{i,q} \sum_{p=1}^{n} (-1)^{p+q} b_p \det\left(A_{\sim p,\sim q}\right)$$
(here, we renamed the summation index k as q in the first sum)
$$= \underbrace{\sum_{q=1}^{n}\sum_{p=1}^{n}}_{\substack{=\sum_{p=1}^{n}\sum_{q=1}^{n} \\ =\sum_{p\in\{1,2,\ldots,n\}}\sum_{q=1}^{n}}} \underbrace{a_{i,q}(-1)^{p+q} b_p}_{=b_p(-1)^{p+q}a_{i,q}} \det\left(A_{\sim p,\sim q}\right)$$
$$= \sum_{p\in\{1,2,\ldots,n\}} \sum_{q=1}^{n} b_p (-1)^{p+q} a_{i,q} \det\left(A_{\sim p,\sim q}\right)$$
$$= \sum_{\substack{p\in\{1,2,\ldots,n\};\\ p\neq i}} \underbrace{\sum_{q=1}^{n} b_p (-1)^{p+q} a_{i,q} \det\left(A_{\sim p,\sim q}\right)}_{\substack{=0 \\ \text{(by (450))}}} + \underbrace{\sum_{q=1}^{n} b_i (-1)^{i+q} a_{i,q} \det\left(A_{\sim i,\sim q}\right)}_{\substack{=\det A\cdot b_i \\ \text{(by (452))}}}$$
(here, we have split off the addend for $p = i$ from the sum)
$$= \underbrace{\sum_{\substack{p\in\{1,2,\ldots,n\};\\ p\neq i}} 0}_{=0} + \det A \cdot b_i = \det A \cdot b_i. \tag{454}$$

Now, let us forget that we fixed i. We thus have proven (454) for every $i \in$

[267]*Proof of (452):* Applying Theorem 6.82 **(a)** to $p = i$, we obtain

$$\det A = \sum_{q=1}^{n} (-1)^{i+q} a_{i,q} \det\left(A_{\sim i,\sim q}\right). \tag{453}$$

Now,

$$\sum_{q=1}^{n} b_i (-1)^{i+q} a_{i,q} \det\left(A_{\sim i,\sim q}\right) = b_i \underbrace{\sum_{q=1}^{n} (-1)^{i+q} a_{i,q} \det\left(A_{\sim i,\sim q}\right)}_{\substack{=\det A \\ \text{(by (453))}}} = b_i \det A = \det A \cdot b_i.$$

This proves (452).

$\{1, 2, \ldots, n\}$. Now, let d be the vector $\left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T$. Thus,

$$d = \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T = \begin{pmatrix} \det\left(A_1^{\#}\right) \\ \det\left(A_2^{\#}\right) \\ \vdots \\ \det\left(A_n^{\#}\right) \end{pmatrix}$$
$$= \left(\det\left(A_i^{\#}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq 1}.$$

The definition of the product of two matrices shows that

$$A \cdot d = \left(\underbrace{\sum_{k=1}^{n} a_{i,k} \det\left(A_k^{\#}\right)}_{\substack{=\det A \cdot b_i \\ \text{(by (454))}}} \right)_{1\leq i\leq n,\ 1\leq j\leq 1}$$
$$\left(\text{since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \text{ and } d = \left(\det\left(A_i^{\#}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq 1}\right)$$
$$= \left(\det A \cdot b_i\right)_{1\leq i\leq n,\ 1\leq j\leq 1} = \left(\det A \cdot b_1, \det A \cdot b_2, \ldots, \det A \cdot b_n\right)^T.$$

Comparing this with

$$\det A \cdot \underbrace{b}_{=(b_1, b_2, \ldots, b_n)^T} = \det A \cdot \left(b_1, b_2, \ldots, b_n\right)^T = \left(\det A \cdot b_1, \det A \cdot b_2, \ldots, \det A \cdot b_n\right)^T,$$

we obtain $A \cdot d = \det A \cdot b$. Since $d = \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T$, we can rewrite this as $A \cdot \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T = \det A \cdot b$. This proves Theorem 6.120 **(a)**.

(b) Theorem 6.110 **(a)** shows that the matrix A is invertible if and only if the element $\det A$ of $\mathbb{K}$ is invertible (in $\mathbb{K}$). Hence, the element $\det A$ of $\mathbb{K}$ is invertible (since the matrix A is invertible). Thus, $\dfrac{1}{\det A}$ is well-defined. Clearly,

$\underbrace{\frac{1}{\det A} \cdot \det A}_{=1} \cdot b = b$, so that

$$b = \frac{1}{\det A} \cdot \underbrace{\det A \cdot b}_{\substack{=A\cdot\left(\det\left(A_1^{\#}\right),\det\left(A_2^{\#}\right),\ldots,\det\left(A_n^{\#}\right)\right)^T \\ \text{(by Theorem 6.120 (a))}}}$$
$$= \frac{1}{\det A} \cdot A \cdot \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T$$
$$= A \cdot \underbrace{\left(\frac{1}{\det A} \cdot \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T\right)}_{\substack{=\left(\frac{1}{\det A}\det\left(A_1^{\#}\right),\frac{1}{\det A}\det\left(A_2^{\#}\right),\ldots,\frac{1}{\det A}\det\left(A_n^{\#}\right)\right)^T \\ =\left(\frac{\det\left(A_1^{\#}\right)}{\det A},\frac{\det\left(A_2^{\#}\right)}{\det A},\ldots,\frac{\det\left(A_n^{\#}\right)}{\det A}\right)^T}}$$
$$= A \cdot \left(\frac{\det\left(A_1^{\#}\right)}{\det A}, \frac{\det\left(A_2^{\#}\right)}{\det A}, \ldots, \frac{\det\left(A_n^{\#}\right)}{\det A}\right)^T.$$

Therefore,

$$A^{-1} \underbrace{b}_{=A\cdot\left(\frac{\det\left(A_1^{\#}\right)}{\det A},\frac{\det\left(A_2^{\#}\right)}{\det A},\ldots,\frac{\det\left(A_n^{\#}\right)}{\det A}\right)^T}$$
$$= \underbrace{A^{-1}A}_{=I_n} \cdot \left(\frac{\det\left(A_1^{\#}\right)}{\det A}, \frac{\det\left(A_2^{\#}\right)}{\det A}, \ldots, \frac{\det\left(A_n^{\#}\right)}{\det A}\right)^T$$
$$= I_n \cdot \left(\frac{\det\left(A_1^{\#}\right)}{\det A}, \frac{\det\left(A_2^{\#}\right)}{\det A}, \ldots, \frac{\det\left(A_n^{\#}\right)}{\det A}\right)^T = \left(\frac{\det\left(A_1^{\#}\right)}{\det A}, \frac{\det\left(A_2^{\#}\right)}{\det A}, \ldots, \frac{\det\left(A_n^{\#}\right)}{\det A}\right)^T.$$

This proves Theorem 6.120 **(b)**. □

6.20. The Desnanot-Jacobi identity

We now move towards more exotic places. In this section[268], we shall prove the *Desnanot-Jacobi identity*, also known as *Lewis Carroll identity*[269]. We will need some

[268]which, unfortunately, has become more technical and tedious than it promised to be – for which I apologize

[269]See [Bresso99, §3.5] for the history of this identity (as well as for a proof different from ours, and for an application). In a nutshell: Desnanot discovered it in 1819; Jacobi proved it in 1833 (and again in 1841); in 1866, Charles Lutwidge Dodgson (better known as Lewis Carroll, although his mathematical works were not printed under this pen name) popularized it by publishing an algorithm for evaluating determinants that made heavy use of this identity.

notations to state this identity in the generality I want; but first I shall state the two best known particular cases (from which the general version can actually be easily derived, although this is not the way I will take):[270]

Proposition 6.122. Let $n \in \mathbb{N}$ be such that $n \geq 2$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n \times n$-matrix. Let A' be the $(n-2) \times (n-2)$-matrix $\left(a_{i+1,j+1}\right)_{1\leq i\leq n-2,\ 1\leq j\leq n-2}$. (In other words, A' is what remains of the matrix A when we remove the first row, the last row, the first column and the last column.) Then,

$$\begin{aligned} &\det A \cdot \det\left(A'\right) \\ &= \det\left(A_{\sim 1,\sim 1}\right) \cdot \det\left(A_{\sim n,\sim n}\right) - \det\left(A_{\sim 1,\sim n}\right) \cdot \det\left(A_{\sim n,\sim 1}\right). \end{aligned}$$

Proposition 6.123. Let $n \in \mathbb{N}$ be such that $n \geq 2$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n \times n$-matrix. Let $\widetilde{A}$ be the $(n-2) \times (n-2)$-matrix $\left(a_{i+2,j+2}\right)_{1\leq i\leq n-2,\ 1\leq j\leq n-2}$. (In other words, $\widetilde{A}$ is what remains of the matrix A when we remove the first two rows and the first two columns.) Then,

$$\begin{aligned} &\det A \cdot \det \widetilde{A} \\ &= \det\left(A_{\sim 1,\sim 1}\right) \cdot \det\left(A_{\sim 2,\sim 2}\right) - \det\left(A_{\sim 1,\sim 2}\right) \cdot \det\left(A_{\sim 2,\sim 1}\right). \end{aligned}$$

Example 6.124. For this example, set $n = 4$ and $A = \begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{pmatrix}$. Then, Proposition 6.122 says that

$$\begin{aligned} &\det \begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{pmatrix} \cdot \det \begin{pmatrix} b_2 & c_2 \\ b_3 & c_3 \end{pmatrix} \\ &= \det \begin{pmatrix} b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \\ b_4 & c_4 & d_4 \end{pmatrix} \cdot \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \\ &\qquad - \det \begin{pmatrix} a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \end{pmatrix} \cdot \det \begin{pmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{pmatrix}. \end{aligned}$$

[270] We shall use the notations of Definition 6.81 throughout this section.

Meanwhile, Proposition 6.123 says that

$$\det \begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{pmatrix} \cdot \det \begin{pmatrix} c_3 & d_3 \\ c_4 & d_4 \end{pmatrix}$$
$$= \det \begin{pmatrix} b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \\ b_4 & c_4 & d_4 \end{pmatrix} \cdot \det \begin{pmatrix} a_1 & c_1 & d_1 \\ a_3 & c_3 & d_3 \\ a_4 & c_4 & d_4 \end{pmatrix}$$
$$- \det \begin{pmatrix} a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \\ a_4 & c_4 & d_4 \end{pmatrix} \cdot \det \begin{pmatrix} b_1 & c_1 & d_1 \\ b_3 & c_3 & d_3 \\ b_4 & c_4 & d_4 \end{pmatrix}.$$

Proposition 6.122 occurs (for instance) in [Zeilbe98, *(Alice)*], in [Bresso99, Theorem 3.12], in [Willia15, Example 3.3] and in [Kratt99, Proposition 10] (without a proof, but with a brief list of applications). Proposition 6.123 occurs (among other places) in [BerBru08, (1)] (with a generalization). The reader can easily see that Proposition 6.123 is equivalent to Proposition 6.122[271]; we shall prove a generalization of both.

Let me now introduce some notations:[272]

Definition 6.125. Let $n \in \mathbb{N}$. Let r and s be two elements of $\{1, 2, \ldots, n\}$ such that $r < s$. Then, $(1, 2, \ldots, \widehat{r}, \ldots, \widehat{s}, \ldots, n)$ will denote the $(n-2)$-tuple

$$\left(\underbrace{1, 2, \ldots, r-1}_{\substack{\text{all integers} \\ \text{from } 1 \text{ to } r-1}}, \underbrace{r+1, r+2, \ldots, s-1}_{\substack{\text{all integers} \\ \text{from } r+1 \text{ to } s-1}}, \underbrace{s+1, s+2, \ldots, n}_{\substack{\text{all integers} \\ \text{from } s+1 \text{ to } n}} \right).$$

In other words, $(1, 2, \ldots, \widehat{r}, \ldots, \widehat{s}, \ldots, n)$ will denote the result of removing the entries r and s from the n-tuple $(1, 2, \ldots, n)$.

We can now state a more general version of the Desnanot-Jacobi identity:

Theorem 6.126. Let $n \in \mathbb{N}$ be such that $n \geq 2$. Let p, q, u and v be four elements of $\{1, 2, \ldots, n\}$ such that $p < q$ and $u < v$. Let A be an $n \times n$-matrix. Then,

$$\det A \cdot \det \left(\operatorname{sub}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} A \right)$$
$$= \det \left(A_{\sim p, \sim u} \right) \cdot \det \left(A_{\sim q, \sim v} \right) - \det \left(A_{\sim p, \sim v} \right) \cdot \det \left(A_{\sim q, \sim u} \right).$$

[271] Indeed, one is easily obtained from the other by switching the 2-nd and the n-th rows of the matrix A and switching the 2-nd and the n-th columns of the matrix A, applying parts **(a)** and **(b)** of Exercise 6.7 and checking that all signs cancel.

[272] Recall that we are using the notations of Definition 6.31, of Definition 6.80, and of Definition 6.81.

Example 6.127. If we set $n = 3$, $p = 1$, $q = 2$, $u = 1$, $v = 3$ and $A = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix}$, then Theorem 6.126 says that

$$\det \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix} \det \begin{pmatrix} c' \end{pmatrix}$$
$$= \det \begin{pmatrix} b' & b'' \\ c' & c'' \end{pmatrix} \cdot \det \begin{pmatrix} a & a' \\ c & c' \end{pmatrix} - \det \begin{pmatrix} b & b' \\ c & c' \end{pmatrix} \cdot \det \begin{pmatrix} a' & a'' \\ c' & c'' \end{pmatrix}.$$

Before we prove this theorem, let me introduce some more notations:

Definition 6.128. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times m}$ be an $n \times m$-matrix.

(a) For every $u \in \{1, 2, \ldots, n\}$, we let $A_{u,\bullet}$ be the u-th row of the matrix A. This $A_{u,\bullet}$ is thus a row vector with m entries, i.e., a $1 \times m$-matrix.

(b) For every $v \in \{1, 2, \ldots, m\}$, we let $A_{\bullet,v}$ be the v-th column of the matrix A. This $A_{\bullet,v}$ is thus a column vector with n entries, i.e., an $n \times 1$-matrix.

(c) For every $u \in \{1, 2, \ldots, n\}$, we set $A_{\sim u,\bullet} = \operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A$. This $A_{\sim u,\bullet}$ is thus an $(n-1) \times m$-matrix. (In more intuitive terms, the definition of $A_{\sim u,\bullet}$ rewrites as follows: $A_{\sim u,\bullet}$ is the matrix obtained from the matrix A by removing the u-th row.)

(d) For every $v \in \{1, 2, \ldots, m\}$, we set $A_{\bullet,\sim v} = \operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,m} A$. This $A_{\bullet,\sim v}$ is thus an $n \times (m-1)$-matrix. (In more intuitive terms, the definition of $A_{\bullet,\sim v}$ rewrites as follows: $A_{\bullet,\sim v}$ is the matrix obtained from the matrix A by removing the v-th column.)

Example 6.129. If $n = 3$, $m = 4$ and $A = \begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{pmatrix}$, then

$$A_{2,\bullet} = \begin{pmatrix} a' & b' & c' & d' \end{pmatrix}, \qquad A_{\bullet,2} = \begin{pmatrix} b \\ b' \\ b'' \end{pmatrix},$$
$$A_{\sim 2,\bullet} = \begin{pmatrix} a & b & c & d \\ a'' & b'' & c'' & d'' \end{pmatrix}, \qquad A_{\bullet,\sim 2} = \begin{pmatrix} a & c & d \\ a' & c' & d' \\ a'' & c'' & d'' \end{pmatrix}.$$

Here are some simple properties of the notations introduced in Definition 6.128:

Proposition 6.130. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times m}$ be an $n \times m$-matrix.

(a) For every $u \in \{1, 2, \ldots, n\}$, we have

$$A_{u,\bullet} = (\text{the } u\text{-th row of the matrix } A) = \operatorname{rows}_u A.$$

(Here, the notation $\operatorname{rows}_u A$ is a particular case of Definition 6.31 **(a)**.)

(b) For every $v \in \{1, 2, \ldots, m\}$, we have

$$A_{\bullet,v} = (\text{the } v\text{-th column of the matrix } A) = \operatorname{cols}_v A.$$

(c) For every $u \in \{1, 2, \ldots, n\}$ and $v \in \{1, 2, \ldots, m\}$, we have

$$(A_{\bullet,\sim v})_{\sim u,\bullet} = (A_{\sim u,\bullet})_{\bullet,\sim v} = A_{\sim u,\sim v}.$$

(d) For every $v \in \{1, 2, \ldots, m\}$ and $w \in \{1, 2, \ldots, v-1\}$, we have $(A_{\bullet,\sim v})_{\bullet,w} = A_{\bullet,w}$.

(e) For every $v \in \{1, 2, \ldots, m\}$ and $w \in \{v, v+1, \ldots, m-1\}$, we have $(A_{\bullet,\sim v})_{\bullet,w} = A_{\bullet,w+1}$.

(f) For every $u \in \{1, 2, \ldots, n\}$ and $w \in \{1, 2, \ldots, u-1\}$, we have $(A_{\sim u,\bullet})_{w,\bullet} = A_{w,\bullet}$.

(g) For every $u \in \{1, 2, \ldots, n\}$ and $w \in \{u, u+1, \ldots, n-1\}$, we have $(A_{\sim u,\bullet})_{w,\bullet} = A_{w+1,\bullet}$.

(h) For every $v \in \{1, 2, \ldots, m\}$ and $w \in \{1, 2, \ldots, v-1\}$, we have

$$(A_{\bullet,\sim v})_{\bullet,\sim w} = \operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A.$$

(i) For every $v \in \{1, 2, \ldots, m\}$ and $w \in \{v, v+1, \ldots, m-1\}$, we have

$$(A_{\bullet,\sim v})_{\bullet,\sim w} = \operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w+1},\ldots,m} A.$$

(j) For every $u \in \{1, 2, \ldots, n\}$ and $w \in \{1, 2, \ldots, u-1\}$, we have

$$(A_{\sim u,\bullet})_{\sim w,\bullet} = \operatorname{rows}_{1,2,\ldots,\widehat{w},\ldots,\widehat{u},\ldots,n} A.$$

(k) For every $u \in \{1, 2, \ldots, n\}$ and $w \in \{u, u+1, \ldots, n-1\}$, we have

$$(A_{\sim u,\bullet})_{\sim w,\bullet} = \operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{w+1},\ldots,n} A.$$

(l) For every $v \in \{1, 2, \ldots, n\}$, $u \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, m\}$ satisfying $u < v$, we have

$$(A_{\sim v,\bullet})_{\sim u,\sim q} = \operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} \left(A_{\bullet,\sim q}\right).$$

Proposition 6.131. Let $n \in \mathbb{N}$. Let u and v be two elements of $\{1, 2, \ldots, n\}$ such that $u < v$. Then, $\left((I_n)_{\bullet,u}\right)_{\sim v,\bullet} = (I_{n-1})_{\bullet,u}$. (Recall that I_m denotes the $m \times m$

identity matrix for each $m \in \mathbb{N}$.)

Exercise 6.39. **(a)** Prove Proposition 6.130 and Proposition 6.131.

(b) Derive Proposition 6.123 and Proposition 6.122 from Theorem 6.126.

Here comes another piece of notation:

Definition 6.132. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times m}$ be an $n \times m$-matrix. Let $v \in \mathbb{K}^{n\times 1}$ be a column vector with n entries. Then, $(A \mid v)$ will denote the $n \times (m+1)$-matrix whose $m+1$ columns are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}, v$ (from left to right). (Informally speaking, $(A \mid v)$ is the matrix obtained when the column vector v is "attached" to A at the right edge.)

Example 6.133. We have $\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid \begin{pmatrix} p \\ q \end{pmatrix}\right) = \begin{pmatrix} a & b & p \\ c & d & q \end{pmatrix}$.

The following properties of the notation introduced in Definition 6.132 are not too hard to see (see the solution of Exercise 6.40 for their proofs), and will be used below:

Proposition 6.134. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times m}$ be an $n \times m$-matrix. Let $v \in \mathbb{K}^{n\times 1}$ be a column vector with n entries.

(a) Every $q \in \{1, 2, \ldots, m\}$ satisfies $(A \mid v)_{\bullet,q} = A_{\bullet,q}$.

(b) We have $(A \mid v)_{\bullet,m+1} = v$.

(c) Every $q \in \{1, 2, \ldots, m\}$ satisfies $(A \mid v)_{\bullet,\sim q} = \left(A_{\bullet,\sim q} \mid v\right)$.

(d) We have $(A \mid v)_{\bullet,\sim(m+1)} = A$.

(e) We have $(A \mid v)_{\sim p,\bullet} = \left(A_{\sim p,\bullet} \mid v_{\sim p,\bullet}\right)$ for every $p \in \{1, 2, \ldots, n\}$.

(f) We have $(A \mid v)_{\sim p,\sim(m+1)} = A_{\sim p,\bullet}$ for every $p \in \{1, 2, \ldots, n\}$.

Proposition 6.135. Let n be a positive integer. Let $A \in \mathbb{K}^{n\times(n-1)}$.

(a) For every $v = (v_1, v_2, \ldots, v_n)^T \in \mathbb{K}^{n\times 1}$, we have

$$\det(A \mid v) = \sum_{i=1}^{n} (-1)^{n+i} v_i \det\left(A_{\sim i,\bullet}\right).$$

(b) For every $p \in \{1, 2, \ldots, n\}$, we have

$$\det\left(A \mid (I_n)_{\bullet,p}\right) = (-1)^{n+p} \det\left(A_{\sim p,\bullet}\right).$$

(Notice that $(I_n)_{\bullet,p}$ is the p-th column of the $n \times n$ identity matrix, i.e., the column vector $\left(\underbrace{0, 0, \ldots, 0}_{p-1 \text{ zeroes}}, 1, \underbrace{0, 0, \ldots, 0}_{n-p \text{ zeroes}}\right)^T$.)

Proposition 6.136. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times n}$.

(a) If $n > 0$, then $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right) = A$.

(b) For every $q \in \{1, 2, \ldots, n\}$, we have $\det\left(A_{\bullet,\sim q} \mid A_{\bullet,q}\right) = (-1)^{n+q} \det A$.

(c) If r and q are two elements of $\{1, 2, \ldots, n\}$ satisfying $r \neq q$, then $\det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) = 0$.

(d) For every $p \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, n\}$, we have $\det\left(A_{\bullet,\sim q} \mid (I_n)_{\bullet,p}\right) = (-1)^{n+p} \det\left(A_{\sim p,\sim q}\right)$.

(e) If u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u < v$, and if r is an element of $\{1, 2, \ldots, n-1\}$ satisfying $r \neq u$, then $\det\left(A_{\bullet,\sim u} \mid (A_{\bullet,\sim v})_{\bullet,r}\right) = 0$.

(f) If u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u < v$, then $(-1)^u \det\left(A_{\bullet,\sim u} \mid (A_{\bullet,\sim v})_{\bullet,u}\right) = (-1)^n \det A$.

Exercise 6.40. Prove Proposition 6.134, Proposition 6.135 and Proposition 6.136.

Now, we can state a slightly more interesting identity:

Proposition 6.137. Let n be a positive integer. Let $A \in \mathbb{K}^{n \times (n-1)}$ and $C \in \mathbb{K}^{n \times n}$. Let $v \in \{1, 2, \ldots, n\}$. Then,

$$\det\left(A_{\sim v,\bullet}\right) \det C = \sum_{q=1}^{n} (-1)^{n+q} \det\left(A \mid C_{\bullet,q}\right) \det\left(C_{\sim v,\sim q}\right).$$

Example 6.138. If we set $n = 3$, $A = \begin{pmatrix} a & a' \\ b & b' \\ c & c' \end{pmatrix}$, $C = \begin{pmatrix} x & x' & x'' \\ y & y' & y'' \\ z & z' & z'' \end{pmatrix}$ and $v = 2$, then Proposition 6.137 states that

$$\begin{aligned} &\det \begin{pmatrix} a & a' \\ c & c' \end{pmatrix} \det \begin{pmatrix} x & x' & x'' \\ y & y' & y'' \\ z & z' & z'' \end{pmatrix} \\ &= \det \begin{pmatrix} a & a' & x \\ b & b' & y \\ c & c' & z \end{pmatrix} \det \begin{pmatrix} x' & x'' \\ z' & z'' \end{pmatrix} - \det \begin{pmatrix} a & a' & x' \\ b & b' & y' \\ c & c' & z' \end{pmatrix} \det \begin{pmatrix} x & x'' \\ z & z'' \end{pmatrix} \\ &\qquad + \det \begin{pmatrix} a & a' & x'' \\ b & b' & y'' \\ c & c' & z'' \end{pmatrix} \det \begin{pmatrix} x & x' \\ z & z' \end{pmatrix}. \end{aligned}$$

Proof of Proposition 6.137. Write the $n \times n$-matrix C in the form $C = \left(c_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Fix $q \in \{1, 2, \ldots, n\}$. Then, $C_{\bullet,q}$ is the q-th column of the matrix C (by the definition of $C_{\bullet,q}$). Thus,

$$\begin{aligned} C_{\bullet,q} &= (\text{the } q\text{-th column of the matrix } C) \\ &= \begin{pmatrix} c_{1,q} \\ c_{2,q} \\ \vdots \\ c_{n,q} \end{pmatrix} \qquad \left(\text{since } C = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ &= \left(c_{1,q}, c_{2,q}, \ldots, c_{n,q}\right)^T. \end{aligned}$$

Now, Proposition 6.135 **(a)** (applied to $C_{\bullet,q}$ and $c_{i,q}$ instead of v and v_i) yields

$$\begin{aligned} \det\left(A \mid C_{\bullet,q}\right) &= \sum_{i=1}^{n} (-1)^{n+i} c_{i,q} \det\left(A_{\sim i,\bullet}\right) \\ &= \sum_{p=1}^{n} (-1)^{n+p} c_{p,q} \det\left(A_{\sim p,\bullet}\right) \end{aligned} \tag{455}$$

(here, we have renamed the summation index i as p).

Now, forget that we fixed q. We thus have proven (455) for each $q \in \{1, 2, \ldots, n\}$.

We have

$$\sum_{q=1}^{n} (-1)^{v+q} c_{v,q} \det\left(C_{\sim v,\sim q}\right) = \det C \tag{456}$$

[273].

On the other hand, every $p \in \{1, 2, \ldots, n\}$ satisfying $p \neq v$ satisfies

$$\sum_{q=1}^{n} (-1)^{p+q} c_{p,q} \det\left(C_{\sim v,\sim q}\right) = 0 \tag{458}$$

[274].

[273] *Proof of (456):* Theorem 6.82 (applied to C, $c_{i,j}$ and v instead of A, $a_{i,j}$ and p) yields

$$\det C = \sum_{q=1}^{n} (-1)^{v+q} c_{v,q} \det\left(C_{\sim v,\sim q}\right). \tag{457}$$

This proves (456).

[274] *Proof of (458):* Let $p \in \{1, 2, \ldots, n\}$ be such that $p \neq v$. Thus, $v \neq p$. Hence, Proposition 6.96 **(a)** (applied to C, $c_{i,j}$, p and v instead of A, $a_{i,j}$, r and p) yields

$$0 = \sum_{q=1}^{n} (-1)^{v+q} c_{p,q} \det\left(C_{\sim v,\sim q}\right). \tag{459}$$

Now, every $q \in \{1, 2, \ldots, n\}$ satisfies

$$(-1)^{p+q} = (-1)^{(p-v)+(v+q)} \qquad (\text{since } p+q = (p-v)+(v+q))$$
$$= (-1)^{p-v} (-1)^{v+q}.$$

Hence,

$$\sum_{q=1}^{n} \underbrace{(-1)^{p+q}}_{=(-1)^{p-v}(-1)^{v+q}} c_{p,q} \det\left(C_{\sim v,\sim q}\right)$$
$$= \sum_{q=1}^{n} (-1)^{p-v} (-1)^{v+q} c_{p,q} \det\left(C_{\sim v,\sim q}\right) = (-1)^{p-v} \underbrace{\sum_{q=1}^{n} (-1)^{v+q} c_{p,q} \det\left(C_{\sim v,\sim q}\right)}_{\substack{=0 \\ (\text{by (459)})}} = 0.$$

This proves (458).

Now,

$$\begin{aligned}
&\sum_{q=1}^{n}(-1)^{n+q}\underbrace{\det\left(A\mid C_{\bullet,q}\right)}_{\substack{=\sum_{p=1}^{n}(-1)^{n+p}c_{p,q}\det\left(A_{\sim p,\bullet}\right)\\ \text{(by (455))}}}\det\left(C_{\sim v,\sim q}\right)\\
&=\sum_{q=1}^{n}(-1)^{n+q}\left(\sum_{p=1}^{n}(-1)^{n+p}c_{p,q}\det\left(A_{\sim p,\bullet}\right)\right)\det\left(C_{\sim v,\sim q}\right)\\
&=\underbrace{\sum_{q=1}^{n}\sum_{p=1}^{n}}_{=\sum_{p=1}^{n}\sum_{q=1}^{n}}\underbrace{(-1)^{n+q}(-1)^{n+p}}_{\substack{=(-1)^{(n+q)+(n+p)}=(-1)^{p+q}\\ \text{(since } (n+q)+(n+p)=2n+p+q\equiv p+q \bmod 2)}}c_{p,q}\det\left(A_{\sim p,\bullet}\right)\det\left(C_{\sim v,\sim q}\right)\\
&=\underbrace{\sum_{p=1}^{n}}_{=\sum_{p\in\{1,2,\ldots,n\}}}\underbrace{\sum_{q=1}^{n}(-1)^{p+q}c_{p,q}\det\left(A_{\sim p,\bullet}\right)\det\left(C_{\sim v,\sim q}\right)}_{=\det\left(A_{\sim p,\bullet}\right)\sum_{q=1}^{n}(-1)^{p+q}c_{p,q}\det\left(C_{\sim v,\sim q}\right)}\\
&=\sum_{p\in\{1,2,\ldots,n\}}\det\left(A_{\sim p,\bullet}\right)\sum_{q=1}^{n}(-1)^{p+q}c_{p,q}\det\left(C_{\sim v,\sim q}\right)\\
&=\det\left(A_{\sim v,\bullet}\right)\underbrace{\sum_{q=1}^{n}(-1)^{v+q}c_{v,q}\det\left(C_{\sim v,\sim q}\right)}_{\substack{=\det C\\ \text{(by (456))}}}\\
&\qquad+\sum_{\substack{p\in\{1,2,\ldots,n\};\\ p\neq v}}\det\left(A_{\sim p,\bullet}\right)\underbrace{\sum_{q=1}^{n}(-1)^{p+q}c_{p,q}\det\left(C_{\sim v,\sim q}\right)}_{\substack{=0\\ \text{(by (458))}}}\\
&\qquad\left(\begin{array}{c}\text{here, we have split off the addend for } p=v \text{ from the sum,}\\ \text{since } v\in\{1,2,\ldots,n\}\end{array}\right)\\
&=\det\left(A_{\sim v,\bullet}\right)\det C+\underbrace{\sum_{\substack{p\in\{1,2,\ldots,n\};\\ p\neq v}}\det\left(A_{\sim p,\bullet}\right)0}_{=0}=\det\left(A_{\sim v,\bullet}\right)\det C.
\end{aligned}$$

This proves Proposition 6.137. □

Now, let us show a purely technical lemma (gathering a few equalities for easy access in a proof further down):

Lemma 6.139. Let n be a positive integer. Let $B \in \mathbb{K}^{n \times (n-1)}$. Let u and v be two elements of $\{1, 2, \ldots, n\}$ such that $u < v$.

Consider the vector $(I_n)_{\bullet,u} \in \mathbb{K}^{n \times 1}$. (This is the u-th column of the identity matrix I_n. [275])

Define an $n \times n$-matrix $C \in \mathbb{K}^{n \times n}$ by

$$C = \left(B \mid (I_n)_{\bullet,u} \right).$$

Then, the following holds:

(a) We have

$$\det \left(C_{\sim v, \sim q} \right) = - (-1)^{n+u} \det \left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} \left(B_{\bullet,\sim q} \right) \right)$$

for every $q \in \{1, 2, \ldots, n-1\}$.

(b) We have

$$(-1)^{n+q} \det \left(C_{\sim v, \sim q} \right) = - (-1)^{q+u} \det \left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} \left(B_{\bullet,\sim q} \right) \right) \tag{460}$$

for every $q \in \{1, 2, \ldots, n-1\}$.

(c) We have

$$C_{\sim v, \sim n} = B_{\sim v, \bullet}. \tag{461}$$

(d) We have

$$C_{\bullet,q} = B_{\bullet,q} \tag{462}$$

for every $q \in \{1, 2, \ldots, n-1\}$.

(e) Any $A \in \mathbb{K}^{n \times (n-1)}$ satisfies

$$\det \left(A \mid C_{\bullet,n} \right) = (-1)^{n+u} \det \left(A_{\sim u, \bullet} \right). \tag{463}$$

(f) We have

$$\det C = (-1)^{n+u} \det \left(B_{\sim u, \bullet} \right). \tag{464}$$

Proof of Lemma 6.139. We have $n - 1 \in \mathbb{N}$ (since n is a positive integer). Also, $u < v \leq n$ (since $v \in \{1, 2, \ldots, n\}$) and thus $u \leq n - 1$ (since u and n are integers). Combining this with $u \geq 1$ (since $u \in \{1, 2, \ldots, n\}$), we obtain $u \in \{1, 2, \ldots, n-1\}$.

(a) Let $q \in \{1, 2, \ldots, n-1\}$. From $C = \left(B \mid (I_n)_{\bullet,u} \right)$, we obtain

$$C_{\bullet,\sim q} = \left(B \mid (I_n)_{\bullet,u} \right)_{\bullet,\sim q} = \left(B_{\bullet,\sim q} \mid (I_n)_{\bullet,u} \right)$$

[275] Explicitly,

$$(I_n)_{\bullet,u} = \Bigl(\underbrace{0, 0, \ldots, 0}_{u-1 \text{ zeroes}}, 1, \underbrace{0, 0, \ldots, 0}_{n-u \text{ zeroes}} \Bigr)^T.$$

(by Proposition 6.134 **(c)**, applied to $n-1$, B and $(I_n)_{\bullet,u}$ instead of m, A and v). But Proposition 6.130 **(c)** (applied to n, n, C, v and q instead of n, m, A, u and v) yields $\left(C_{\bullet,\sim q}\right)_{\sim v,\bullet} = \left(C_{\sim v,\bullet}\right)_{\bullet,\sim q} = C_{\sim v,\sim q}$. Hence,

$$C_{\sim v,\sim q} = \left(\underbrace{C_{\bullet,\sim q}}_{=\left(B_{\bullet,\sim q} \mid (I_n)_{\bullet,u}\right)}\right)_{\sim v,\bullet} = \left(B_{\bullet,\sim q} \mid (I_n)_{\bullet,u}\right)_{\sim v,\bullet}$$
$$= \left(\left(B_{\bullet,\sim q}\right)_{\sim v,\bullet} \mid \left((I_n)_{\bullet,u}\right)_{\sim v,\bullet}\right) \qquad (465)$$

(by Proposition 6.134 **(e)**, applied to n, $n-2$, $B_{\bullet,\sim q}$, $(I_n)_{\bullet,u}$ and v instead of n, m, A, v and p).

On the other hand, Proposition 6.130 **(c)** (applied to n, $n-1$, B, v and q instead of n, m, A, u and v) yields $\left(B_{\bullet,\sim q}\right)_{\sim v,\bullet} = \left(B_{\sim v,\bullet}\right)_{\bullet,\sim q} = B_{\sim v,\sim q}$. Now, (465) becomes

$$C_{\sim v,\sim q} = \left(\underbrace{\left(B_{\bullet,\sim q}\right)_{\sim v,\bullet}}_{=\left(B_{\sim v,\bullet}\right)_{\bullet,\sim q}} \mid \underbrace{\left((I_n)_{\bullet,u}\right)_{\sim v,\bullet}}_{\substack{=(I_{n-1})_{\bullet,u} \\ \text{(by Proposition 6.131)}}}\right) = \left(\left(B_{\sim v,\bullet}\right)_{\bullet,\sim q} \mid (I_{n-1})_{\bullet,u}\right).$$

Hence,

$$\det \underbrace{\left(C_{\sim v,\sim q}\right)}_{=\left(\left(B_{\sim v,\bullet}\right)_{\bullet,\sim q} \mid (I_{n-1})_{\bullet,u}\right)} = \det\left(\left(B_{\sim v,\bullet}\right)_{\bullet,\sim q} \mid (I_{n-1})_{\bullet,u}\right)$$
$$= (-1)^{(n-1)+u} \det\left(\left(B_{\sim v,\bullet}\right)_{\sim u,\sim q}\right)$$

(by Proposition 6.136 **(d)**, applied to $n-1$, $B_{\sim v,\bullet}$ and u instead of n, A and p). Thus,

$$\det\left(C_{\sim v,\sim q}\right) = \underbrace{(-1)^{(n-1)+u}}_{\substack{=(-1)^{n+u+1} \\ \text{(since } (n-1)+u=(n+u+1)-2 \\ \equiv n+u+1 \bmod 2)}} \det\left(\underbrace{\left(B_{\sim v,\bullet}\right)_{\sim u,\sim q}}_{\substack{=\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right) \\ \text{(by Proposition 6.130 (l),} \\ \text{applied to } B \text{ and } n-1 \text{ instead of } A \text{ and } m)}}\right)$$
$$= \underbrace{(-1)^{n+u+1}}_{=-(-1)^{n+u}} \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)$$
$$= -(-1)^{n+u} \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right).$$

This proves Lemma 6.139 **(a)**.

(b) Let $q \in \{1, 2, \ldots, n-1\}$. Then,

$$\begin{aligned}
&(-1)^{n+q} \underbrace{\det\left(C_{\sim v,\sim q}\right)}_{\substack{=-(-1)^{n+u}\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) \\ \text{(by Lemma 6.139 (a))}}} \\
&= (-1)^{n+q}\left(-(-1)^{n+u}\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\right) \\
&= - \underbrace{(-1)^{n+q}(-1)^{n+u}}_{\substack{=(-1)^{(n+q)+(n+u)}=(-1)^{q+u} \\ \text{(since } (n+q)+(n+u)=2n+q+u\equiv q+u \bmod 2)}} \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) \\
&= -(-1)^{q+u}\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right).
\end{aligned}$$

This proves Lemma 6.139 **(b)**.

(c) We have $C = \left(B \mid (I_n)_{\bullet,u}\right)$. Thus,

$$\begin{aligned}
C_{\sim v,\sim n} &= \left(B \mid (I_n)_{\bullet,u}\right)_{\sim v,\sim n} = \left(B \mid (I_n)_{\bullet,u}\right)_{\sim v,\sim((n-1)+1)} \qquad (\text{since } n = (n-1)+1) \\
&= B_{\sim v,\bullet} \qquad \left(\begin{array}{c}\text{by Proposition 6.134 (f), applied to} \\ n-1,\ B \text{ and } (I_n)_{\bullet,u} \text{ instead of } m,\ A \text{ and } v\end{array}\right).
\end{aligned}$$

This proves Lemma 6.139 **(c)**.

(d) Let $q \in \{1, 2, \ldots, n-1\}$. From $C = \left(B \mid (I_n)_{\bullet,u}\right)$, we obtain

$$C_{\bullet,q} = \left(B \mid (I_n)_{\bullet,u}\right)_{\bullet,q} = B_{\bullet,q}$$

(by Proposition 6.134 **(a)**, applied to $n-1$, B and $(I_n)_{\bullet,u}$ instead of m, A and v). This proves Lemma 6.139 **(d)**.

(e) Let $A \in \mathbb{K}^{(n-1)\times n}$. Proposition 6.134 **(b)** (applied to $n-1$, B and $(I_n)_{\bullet,u}$ instead of m, A and v) yields $\left(B \mid (I_n)_{\bullet,u}\right)_{\bullet,(n-1)+1} = (I_n)_{\bullet,u}$. This rewrites as $\left(B \mid (I_n)_{\bullet,u}\right)_{\bullet,n} = (I_n)_{\bullet,u}$ (since $(n-1)+1 = n$). Now, $C = \left(B \mid (I_n)_{\bullet,u}\right)$, so that

$$C_{\bullet,n} = \left(B \mid (I_n)_{\bullet,u}\right)_{\bullet,n} = (I_n)_{\bullet,u}.$$

Hence,

$$\det\left(A \mid \underbrace{C_{\bullet,n}}_{=(I_n)_{\bullet,u}}\right) = \det\left(A \mid (I_n)_{\bullet,u}\right) = (-1)^{n+u}\det\left(A_{\sim u,\bullet}\right)$$

(by Proposition 6.135 **(b)**, applied to $p = u$). This proves Lemma 6.139 **(e)**.

(f) We have

$$\det \underbrace{C}_{=\left(B|(I_n)_{\bullet,u}\right)} = \det\left(B \mid (I_n)_{\bullet,u}\right) = (-1)^{n+u} \det\left(B_{\sim u,\bullet}\right)$$

(by Proposition 6.135 **(b)**, applied to B and u instead of A and p). This proves Lemma 6.139 **(f)**. ☐

Next, we claim the following:

Proposition 6.140. Let n be a positive integer. Let $A \in \mathbb{K}^{n\times(n-1)}$ and $B \in \mathbb{K}^{n\times(n-1)}$. Let u and v be two elements of $\{1, 2, \ldots, n\}$ such that $u < v$. Then,

$$\sum_{r=1}^{n-1} (-1)^r \det\left(A \mid B_{\bullet,r}\right) \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim r}\right)\right)$$
$$= (-1)^n \left(\det\left(A_{\sim u,\bullet}\right) \det\left(B_{\sim v,\bullet}\right) - \det\left(A_{\sim v,\bullet}\right) \det\left(B_{\sim u,\bullet}\right)\right).$$

Example 6.141. If we set $n = 3$, $A = \begin{pmatrix} a & a' \\ b & b' \\ c & c' \end{pmatrix}$, $B = \begin{pmatrix} x & x' \\ y & y' \\ z & z' \end{pmatrix}$, $u = 1$ and $v = 2$, then Proposition 6.140 claims that

$$-\det\begin{pmatrix} a & a' & x \\ b & b' & y \\ c & c' & z \end{pmatrix} \det\begin{pmatrix} z' \end{pmatrix} + \det\begin{pmatrix} a & a' & x' \\ b & b' & y' \\ c & c' & z' \end{pmatrix} \det\begin{pmatrix} z \end{pmatrix}$$
$$= (-1)^3 \left(\det\begin{pmatrix} b & b' \\ c & c' \end{pmatrix} \det\begin{pmatrix} x & x' \\ z & z' \end{pmatrix} - \det\begin{pmatrix} a & a' \\ c & c' \end{pmatrix} \det\begin{pmatrix} y & y' \\ z & z' \end{pmatrix}\right).$$

Proof of Proposition 6.140. We have $n - 1 \in \mathbb{N}$ (since n is a positive integer).

Define an $n \times n$-matrix $C \in \mathbb{K}^{n\times n}$ as in Lemma 6.139.

Proposition 6.137 yields

$$\begin{aligned}
&\det\left(A_{\sim v,\bullet}\right)\det C \\
&= \sum_{q=1}^{n} \underbrace{(-1)^{n+q}\det\left(A \mid C_{\bullet,q}\right)}_{=\det\left(A\mid C_{\bullet,q}\right)(-1)^{n+q}} \det\left(C_{\sim v,\sim q}\right) = \sum_{q=1}^{n} \det\left(A \mid C_{\bullet,q}\right)(-1)^{n+q}\det\left(C_{\sim v,\sim q}\right) \\
&= \sum_{q=1}^{n-1} \det\left(A \mid \underbrace{C_{\bullet,q}}_{\substack{=B_{\bullet,q} \\ \text{(by (462))}}}\right) \underbrace{(-1)^{n+q}\det\left(C_{\sim v,\sim q}\right)}_{\substack{=-(-1)^{q+u}\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) \\ \text{(by (460))}}} \\
&\qquad + \underbrace{\det\left(A \mid C_{\bullet,n}\right)}_{\substack{=(-1)^{n+u}\det\left(A_{\sim u,\bullet}\right) \\ \text{(by (463))}}} \underbrace{(-1)^{n+n}}_{\substack{=1 \\ \text{(since } n+n=2n \text{ is even)}}} \det\left(\underbrace{C_{\sim v,\sim n}}_{\substack{=B_{\sim v,\bullet} \\ \text{(by (461))}}}\right) \\
&\qquad \left(\text{here, we have split off the addend for } q=n \text{ from the sum}\right) \\
&= \underbrace{\sum_{q=1}^{n-1} \det\left(A \mid B_{\bullet,q}\right)\left(-(-1)^{q+u}\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\right)}_{=-\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A\mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)} \\
&\qquad + (-1)^{n+u}\det\left(A_{\sim u,\bullet}\right)\det\left(B_{\sim v,\bullet}\right) \\
&= -\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A \mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) \\
&\qquad + (-1)^{n+u}\det\left(A_{\sim u,\bullet}\right)\det\left(B_{\sim v,\bullet}\right).
\end{aligned}$$

Adding $\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A \mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)$ to both sides of this equality, we obtain

$$\begin{aligned}
&\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A \mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) + \det\left(A_{\sim v,\bullet}\right)\det C \\
&= (-1)^{n+u}\det\left(A_{\sim u,\bullet}\right)\det\left(B_{\sim v,\bullet}\right).
\end{aligned}$$

Subtracting $\det(A_{\sim v,\bullet})\det C$ from both sides of this equality, we find

$$\begin{aligned}
&\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A\mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\\
&=(-1)^{n+u}\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-\det(A_{\sim v,\bullet})\underbrace{\det C}_{\substack{=(-1)^{n+u}\det(B_{\sim u,\bullet})\\ \text{(by (464))}}}\\
&=(-1)^{n+u}\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-\underbrace{\det(A_{\sim v,\bullet})(-1)^{n+u}}_{=(-1)^{n+u}\det(A_{\sim v,\bullet})}\det(B_{\sim u,\bullet})\\
&=(-1)^{n+u}\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-(-1)^{n+u}\det(A_{\sim v,\bullet})\det(B_{\sim u,\bullet})\\
&=(-1)^{n+u}\left(\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-\det(A_{\sim v,\bullet})\det(B_{\sim u,\bullet})\right).
\end{aligned}$$

Multiplying both sides of this equality by $(-1)^u$, we obtain

$$\begin{aligned}
&(-1)^u\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A\mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\\
&=\underbrace{(-1)^u(-1)^{n+u}}_{\substack{=(-1)^{u+(n+u)}=(-1)^n\\ \text{(since } u+(n+u)=2u+n\equiv n\bmod 2)}}\left(\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-\det(A_{\sim v,\bullet})\det(B_{\sim u,\bullet})\right)\\
&=(-1)^n\left(\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-\det(A_{\sim v,\bullet})\det(B_{\sim u,\bullet})\right),
\end{aligned}$$

so that

$$\begin{aligned}
&(-1)^n\left(\det(A_{\sim u,\bullet})\det(B_{\sim v,\bullet})-\det(A_{\sim v,\bullet})\det(B_{\sim u,\bullet})\right)\\
&=(-1)^u\sum_{q=1}^{n-1}(-1)^{q+u}\det\left(A\mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\\
&=\sum_{q=1}^{n-1}\underbrace{(-1)^u(-1)^{q+u}}_{\substack{=(-1)^{u+(q+u)}=(-1)^q\\ \text{(since } u+(q+u)=2u+q\equiv q\bmod 2)}}\det\left(A\mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\\
&=\sum_{q=1}^{n-1}(-1)^q\det\left(A\mid B_{\bullet,q}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)\\
&=\sum_{r=1}^{n-1}(-1)^r\det\left(A\mid B_{\bullet,r}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim r}\right)\right)
\end{aligned}$$

(here, we have renamed the summation index q as r). This proves Proposition 6.140. □

Now, we can finally prove Theorem 6.126:

Proof of Theorem 6.126. We have $v \in \{1, 2, \ldots, n\}$ and thus $v \leq n$. Hence, $u < v \leq n$, so that $u \leq n-1$ (since u and n are integers). Also, $u \in \{1, 2, \ldots, n\}$, so that $1 \leq u$. Combining $1 \leq u$ with $u \leq n-1$, we obtain $u \in \{1, 2, \ldots, n-1\}$.

Proposition 6.79 **(d)** (applied to n, $n-2$, $(1, 2, \ldots, \widehat{p}, \ldots, \widehat{q}, \ldots, n)$, $n-2$ and $(1, 2, \ldots, \widehat{u}, \ldots, \widehat{v}, \ldots, n)$ instead of m, u, $(i_1, i_2, \ldots, i_u)$, v and $(j_1, j_2, \ldots, j_v)$) yields

$$\begin{aligned} \operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A &= \operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} \left(\operatorname{cols}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} A\right) \\ &= \operatorname{cols}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} \left(\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A\right). \end{aligned} \tag{466}$$

We have $u < v$, so that $u \leq v-1$ (since u and v are integers). Combining this with $1 \leq u$, we obtain $u \in \{1, 2, \ldots, v-1\}$. Thus, Proposition 6.130 **(h)** (applied to $m = n$ and $w = u$) yields $(A_{\bullet,\sim v})_{\bullet,\sim u} = \operatorname{cols}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} A$. Thus,

$$\begin{aligned} &\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} \left(\underbrace{(A_{\bullet,\sim v})_{\bullet,\sim u}}_{=\operatorname{cols}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} A} \right) \\ &= \operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} \left(\operatorname{cols}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} A\right) \\ &= \operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A \qquad (\text{by (466)}). \end{aligned} \tag{467}$$

Proposition 6.130 **(c)** (applied to n, p and u instead of m, u and v) yields $(A_{\bullet,\sim u})_{\sim p,\bullet} = (A_{\sim p,\bullet})_{\bullet,\sim u} = A_{\sim p,\sim u}$. Similarly, $(A_{\bullet,\sim v})_{\sim p,\bullet} = (A_{\sim p,\bullet})_{\bullet,\sim v} = A_{\sim p,\sim v}$ and $(A_{\bullet,\sim u})_{\sim q,\bullet} = (A_{\sim q,\bullet})_{\bullet,\sim u} = A_{\sim q,\sim u}$ and $(A_{\bullet,\sim v})_{\sim q,\bullet} = (A_{\sim q,\bullet})_{\bullet,\sim v} = A_{\sim q,\sim v}$.

The integer n is positive (since $n \geq 2$). Thus, Proposition 6.140 (applied to $A_{\bullet,\sim u}$, $A_{\bullet,\sim v}$, p and q instead of A, B, u and v) yields

$$\begin{aligned} &\sum_{r=1}^{n-1} (-1)^r \det\left(A_{\bullet,\sim u} \mid (A_{\bullet,\sim v})_{\bullet,r}\right) \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}\left((A_{\bullet,\sim v})_{\bullet,\sim r}\right)\right) \\ &= (-1)^n \left(\det\left(\underbrace{(A_{\bullet,\sim u})_{\sim p,\bullet}}_{=A_{\sim p,\sim u}}\right) \det\left(\underbrace{(A_{\bullet,\sim v})_{\sim q,\bullet}}_{=A_{\sim q,\sim v}}\right) \right. \\ &\qquad \left. - \det\left(\underbrace{(A_{\bullet,\sim u})_{\sim q,\bullet}}_{=A_{\sim q,\sim u}}\right) \det\left(\underbrace{(A_{\bullet,\sim v})_{\sim p,\bullet}}_{=A_{\sim p,\sim v}}\right) \right) \\ &= (-1)^n \left(\det\left(A_{\sim p,\sim u}\right) \det\left(A_{\sim q,\sim v}\right) - \det\left(A_{\sim q,\sim u}\right) \det\left(A_{\sim p,\sim v}\right)\right). \end{aligned}$$

Hence,

$$
\begin{aligned}
&(-1)^n \left(\det\left(A_{\sim p,\sim u}\right)\det\left(A_{\sim q,\sim v}\right) - \det\left(A_{\sim q,\sim u}\right)\det\left(A_{\sim p,\sim v}\right)\right) \\
&= \underbrace{\sum_{r=1}^{n-1}}_{=\sum_{r\in\{1,2,\ldots,n-1\}}} (-1)^r \det\left(A_{\bullet,\sim u} \mid \left(A_{\bullet,\sim v}\right)_{\bullet,r}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}\left(\left(A_{\bullet,\sim v}\right)_{\bullet,\sim r}\right)\right) \\
&= \sum_{r\in\{1,2,\ldots,n-1\}} (-1)^r \det\left(A_{\bullet,\sim u} \mid \left(A_{\bullet,\sim v}\right)_{\bullet,r}\right)\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}\left(\left(A_{\bullet,\sim v}\right)_{\bullet,\sim r}\right)\right) \\
&= \sum_{\substack{r\in\{1,2,\ldots,n-1\};\\ r\neq u}} (-1)^r \underbrace{\det\left(A_{\bullet,\sim u} \mid \left(A_{\bullet,\sim v}\right)_{\bullet,r}\right)}_{\substack{=0\\ \text{(by Proposition 6.136 (e))}}}\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}\left(\left(A_{\bullet,\sim v}\right)_{\bullet,\sim r}\right)\right) \\
&\qquad + \underbrace{(-1)^u \det\left(A_{\bullet,\sim u} \mid \left(A_{\bullet,\sim v}\right)_{\bullet,u}\right)}_{\substack{=(-1)^n \det A\\ \text{(by Proposition 6.136 (f))}}}\det\left(\underbrace{\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}\left(\left(A_{\bullet,\sim v}\right)_{\bullet,\sim u}\right)}_{\substack{=\operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A\\ \text{(by (467))}}}\right) \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } r=u \text{ from the sum,}\\ \text{since } u\in\{1,2,\ldots,n-1\}\end{array}\right) \\
&= \underbrace{\sum_{\substack{r\in\{1,2,\ldots,n-1\};\\ r\neq u}} (-1)^r\, 0 \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}\left(\left(A_{\bullet,\sim v}\right)_{\bullet,\sim r}\right)\right)}_{=0} \\
&\qquad + (-1)^n \det A \cdot \det\left(\operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A\right) \\
&= (-1)^n \det A \cdot \det\left(\operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A\right).
\end{aligned}
$$

Multiplying both sides of this equality by $(-1)^n$, we obtain

$$
\begin{aligned}
&(-1)^n (-1)^n \left(\det\left(A_{\sim p,\sim u}\right)\det\left(A_{\sim q,\sim v}\right) - \det\left(A_{\sim q,\sim u}\right)\det\left(A_{\sim p,\sim v}\right)\right) \\
&= \underbrace{(-1)^n (-1)^n}_{\substack{=(-1)^{n+n}=1\\ \text{(since } n+n=2n \text{ is even)}}} \det A \cdot \det\left(\operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A\right) \\
&= \det A \cdot \det\left(\operatorname{sub}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n} A\right).
\end{aligned}
$$

Thus,

$$\begin{aligned}
&\det A \cdot \det\left(\operatorname{sub}_{1,2,\ldots,\widehat{p},\ldots,\widehat{q},\ldots,n}^{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n} A\right) \\
&= \underbrace{(-1)^n (-1)^n}_{\substack{=(-1)^{n+n}=1 \\ \text{(since } n+n=2n \text{ is even)}}} \left(\det\left(A_{\sim p,\sim u}\right)\det\left(A_{\sim q,\sim v}\right) - \det\left(A_{\sim q,\sim u}\right)\det\left(A_{\sim p,\sim v}\right)\right) \\
&= \det\left(A_{\sim p,\sim u}\right)\det\left(A_{\sim q,\sim v}\right) - \det\left(A_{\sim q,\sim u}\right)\det\left(A_{\sim p,\sim v}\right).
\end{aligned}$$

This proves Theorem 6.126. □

Now that Theorem 6.126 is proven, we conclude that Proposition 6.123 and Proposition 6.122 hold as well (because in Exercise 6.39 **(b)**, these two propositions have been derived from Theorem 6.126).

Exercise 6.41. Let n be a positive integer. Let $B \in \mathbb{K}^{n\times(n-1)}$. Fix $q \in \{1,2,\ldots,n-1\}$. For every $x \in \{1,2,\ldots,n\}$, set

$$\alpha_x = \det\left(B_{\sim x,\bullet}\right).$$

For every two elements x and y of $\{1,2,\ldots,n\}$ satisfying $x < y$, set

$$\beta_{x,y} = \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{x},\ldots,\widehat{y},\ldots,n}\left(B_{\bullet,\sim q}\right)\right).$$

(Note that this depends on q, but we do not mention q in the notation because q is fixed.)

Let u, v and w be three elements of $\{1,2,\ldots,n\}$ such that $u < v < w$. Thus, $\beta_{u,v}$, $\beta_{v,w}$ and $\beta_{u,w}$ are well-defined elements of $\mathbb{K}$. Prove that

$$\alpha_u \beta_{v,w} + \alpha_w \beta_{u,v} = \alpha_v \beta_{u,w}.$$

Example 6.142. If we set $n = 4$, $B = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix}$, $q = 3$, $u = 1$, $v = 2$ and $w = 3$, then Exercise 6.41 says that

$$\begin{aligned}
&\det\begin{pmatrix} b & b' & b'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix} \cdot \det\begin{pmatrix} a & a' \\ d & d' \end{pmatrix} + \det\begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ d & d' & d'' \end{pmatrix} \cdot \det\begin{pmatrix} c & c' \\ d & d' \end{pmatrix} \\
&= \det\begin{pmatrix} a & a' & a'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix} \cdot \det\begin{pmatrix} b & b' \\ d & d' \end{pmatrix}.
\end{aligned}$$

Remark 6.143. Exercise 6.41 appears in [KenWil14, proof of Theorem 9], where it (or, rather, a certain transformation of determinant expressions that relies on it) is called the "jaw move".

Exercise 6.42. Let $n \in \mathbb{N}$. Let A be an alternating $n \times n$-matrix. (See Definition 6.72 **(b)** for what this means.) Let S be any $n \times n$-matrix. Prove that each entry of the matrix $(\operatorname{adj} S)^T A (\operatorname{adj} S)$ is a multiple of $\det S$.

6.21. The Plücker relation

The following section is devoted to the *Plücker relations*, or, rather, one of the many things that tend to carry this name in the literature[276]. The proofs will be fairly short, since we did much of the necessary work in Section 6.20 already.

We shall use the notations of Definition 6.128 throughout this section.

We begin with the following identity:

Proposition 6.144. Let n be a positive integer. Let $B \in \mathbb{K}^{n \times (n-1)}$. Then:

(a) We have

$$\sum_{r=1}^{n} (-1)^r \det (B_{\sim r, \bullet}) B_{r, \bullet} = 0_{1 \times (n-1)}.$$

(Recall that the product $\det (B_{\sim r, \bullet}) B_{r, \bullet}$ in this equality is the product of the scalar $\det (B_{\sim r, \bullet}) \in \mathbb{K}$ with the row vector $B_{r, \bullet} \in \mathbb{K}^{1 \times (n-1)}$; as all such products, it is computed entrywise, i.e., by the formula $\lambda (a_1, a_2, \ldots, a_{n-1}) = (\lambda a_1, \lambda a_2, \ldots, \lambda a_{n-1})$.)

(b) Write the matrix B in the form $B = (b_{i,j})_{1 \leq i \leq n, \ 1 \leq j \leq n-1}$. Then,

$$\sum_{r=1}^{n} (-1)^r \det (B_{\sim r, \bullet}) b_{r,q} = 0$$

for every $q \in \{1, 2, \ldots, n-1\}$.

Note that part **(a)** of Proposition 6.144 claims an equality between two row vectors (indeed, $B_{r, \bullet}$ is a row vector with $n-1$ entries for each $r \in \{1, 2, \ldots, n\}$), whereas part **(b)** claims an equality between two elements of $\mathbb{K}$ (for each $q \in \{1, 2, \ldots, n-1\}$). That said, the two parts are essentially restatements of one another, and we will derive part **(a)** from part **(b)** soon enough. Let us first illustrate Proposition 6.144 on an example:

[276]Most of the relevant literature, unfortunately, is not very elementary, as the Plücker relations are at their most useful in the algebraic geometry of the Grassmannian and of flag varieties ("Schubert calculus"). See [KleLak72], [Jacobs10, §3.4] and [Fulton97, §9.1] for expositions (all three, however, well above the level of the present notes).

Example 6.145. For this example, set $n = 4$ and $B = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix}$. Then, Proposition 6.144 **(a)** says that

$$-\det\begin{pmatrix} b & b' & b'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix} \cdot \begin{pmatrix} a & a' & a'' \end{pmatrix} + \det\begin{pmatrix} a & a' & a'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix} \cdot \begin{pmatrix} b & b' & b'' \end{pmatrix}$$
$$- \det\begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ d & d' & d'' \end{pmatrix} \cdot \begin{pmatrix} c & c' & c'' \end{pmatrix} + \det\begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix} \cdot \begin{pmatrix} d & d' & d'' \end{pmatrix}$$
$$= 0_{1\times 3}.$$

Proposition 6.144 **(b)** (applied to $q = 3$) yields

$$-\det\begin{pmatrix} b & b' & b'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix} \cdot a'' + \det\begin{pmatrix} a & a' & a'' \\ c & c' & c'' \\ d & d' & d'' \end{pmatrix} \cdot b''$$
$$- \det\begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ d & d' & d'' \end{pmatrix} \cdot c'' + \det\begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix} \cdot d''$$
$$= 0.$$

Proof of Proposition 6.144. **(b)** Let $q \in \{1, 2, \ldots, n-1\}$.

We shall use the notation introduced in Definition 6.132. We know that B is an $n \times (n-1)$-matrix (since $B \in \mathbb{K}^{n\times(n-1)}$). Thus, $\left(B \mid B_{\bullet,q}\right)$ is an $n \times n$-matrix. This $n \times n$-matrix $\left(B \mid B_{\bullet,q}\right)$ is defined as the $n \times ((n-1)+1)$-matrix whose columns are $B_{\bullet,1}, B_{\bullet,2}, \ldots, B_{\bullet,n-1}, B_{\bullet,q}$; thus, it has two equal columns (indeed, the column vector $B_{\bullet,q}$ appears twice among the columns $B_{\bullet,1}, B_{\bullet,2}, \ldots, B_{\bullet,n-1}, B_{\bullet,q}$). Thus, Exercise 6.7 **(f)** (applied to $A = \left(B \mid B_{\bullet,q}\right)$) shows that $\det\left(B \mid B_{\bullet,q}\right) = 0$.

But $B_{\bullet,q}$ is the q-th column of the matrix B (by the definition of $B_{\bullet,q}$). Thus,

$$\begin{aligned} B_{\bullet,q} &= (\text{the } q\text{-th column of the matrix } B) \\ &= \begin{pmatrix} b_{1,q} \\ b_{2,q} \\ \vdots \\ b_{n,q} \end{pmatrix} \qquad \left(\text{since } B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n-1}\right) \\ &= \left(b_{1,q}, b_{2,q}, \ldots, b_{n,q}\right)^T. \end{aligned}$$

Hence, Proposition 6.135 **(a)** (applied to B, $B_{\bullet,q}$ and $b_{i,q}$ instead of A, v and v_i) yields

$$\det\left(B \mid B_{\bullet,q}\right) = \sum_{i=1}^{n} (-1)^{n+i} b_{i,q} \det\left(B_{\sim i,\bullet}\right).$$

Comparing this with $\det\left(B \mid B_{\bullet,q}\right) = 0$, we obtain

$$0 = \sum_{i=1}^{n} (-1)^{n+i} b_{i,q} \det\left(B_{\sim i,\bullet}\right).$$

Multiplying both sides of this equality by $(-1)^n$, we obtain

$$\begin{aligned} 0 &= (-1)^n \sum_{i=1}^{n} (-1)^{n+i} b_{i,q} \det\left(B_{\sim i,\bullet}\right) \\ &= \sum_{i=1}^{n} \underbrace{(-1)^n (-1)^{n+i}}_{\substack{=(-1)^{n+(n+i)}=(-1)^i \\ (\text{since } n+(n+i)=2n+i\equiv i \bmod 2)}} \underbrace{b_{i,q} \det\left(B_{\sim i,\bullet}\right)}_{=\det\left(B_{\sim i,\bullet}\right) b_{i,q}} \\ &= \sum_{i=1}^{n} (-1)^i \det\left(B_{\sim i,\bullet}\right) b_{i,q} = \sum_{r=1}^{n} (-1)^r \det\left(B_{\sim r,\bullet}\right) b_{r,q} \end{aligned}$$

(here, we renamed the summation index i as r). This proves Proposition 6.144 **(b)**.

(a) Write the matrix B in the form $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n-1}$. For every $r \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned} B_{r,\bullet} &= (\text{the } r\text{-th row of the matrix } B) \\ &\qquad \left(\begin{array}{c}\text{since } B_{r,\bullet} \text{ is the } r\text{-th row of the matrix } B \\ (\text{by the definition of } B_{r,\bullet})\end{array}\right) \\ &= \left(b_{r,1}, b_{r,2}, \ldots, b_{r,n-1}\right) \qquad \left(\text{since } B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n-1}\right) \\ &= \left(b_{r,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq n-1}. \end{aligned}$$

Thus,

$$\begin{aligned} &\sum_{r=1}^{n} (-1)^r \det\left(B_{\sim r,\bullet}\right) \underbrace{B_{r,\bullet}}_{=\left(b_{r,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq n-1}} \\ &= \sum_{r=1}^{n} \underbrace{(-1)^r \det\left(B_{\sim r,\bullet}\right) \left(b_{r,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq n-1}}_{=\left((-1)^r \det\left(B_{\sim r,\bullet}\right) b_{r,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq n-1}} = \sum_{r=1}^{n} \left((-1)^r \det\left(B_{\sim r,\bullet}\right) b_{r,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq n-1} \\ &= \left(\underbrace{\sum_{r=1}^{n} (-1)^r \det\left(B_{\sim r,\bullet}\right) b_{r,j}}_{\substack{=0 \\ (\text{by Proposition 6.144 } \textbf{(b)}, \text{ applied to } q=j)}}\right)_{1\leq i\leq 1,\ 1\leq j\leq n-1} = (0)_{1\leq i\leq 1,\ 1\leq j\leq n-1} = 0_{1\times(n-1)}. \end{aligned}$$

This proves Proposition 6.144 **(a)**. □

Let us next state a variant of Proposition 6.144 where rows are replaced by columns (and n is renamed as $n+1$):

Proposition 6.146. Let $n \in \mathbb{N}$. Let $B \in \mathbb{K}^{n\times(n+1)}$. Then:

(a) We have

$$\sum_{r=1}^{n+1} (-1)^r \det(B_{\bullet,\sim r}) B_{\bullet,r} = 0_{n\times 1}.$$

(b) Write the matrix B in the form $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n+1}$. Then,

$$\sum_{r=1}^{n+1} (-1)^r \det(B_{\bullet,\sim r}) b_{q,r} = 0$$

for every $q \in \{1,2,\ldots,n\}$.

Example 6.147. For this example, set $n = 2$ and $B = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \end{pmatrix}$. Then, Proposition 6.146 **(a)** says that

$$-\det\begin{pmatrix} a' & a'' \\ b' & b'' \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} + \det\begin{pmatrix} a & a'' \\ b & b'' \end{pmatrix} \cdot \begin{pmatrix} a' \\ b' \end{pmatrix} - \det\begin{pmatrix} a & a' \\ b & b' \end{pmatrix} \cdot \begin{pmatrix} a'' \\ b'' \end{pmatrix}$$
$$= 0_{2\times 1}.$$

Proposition 6.146 **(b)** (applied to $q = 2$) yields

$$-\det\begin{pmatrix} a' & a'' \\ b' & b'' \end{pmatrix} \cdot b + \det\begin{pmatrix} a & a'' \\ b & b'' \end{pmatrix} \cdot b' - \det\begin{pmatrix} a & a' \\ b & b' \end{pmatrix} \cdot b'' = 0.$$

We shall obtain Proposition 6.146 by applying Proposition 6.144 to $n+1$ and B^T [277] instead of n and B. For this, we shall need a really simple lemma:

Lemma 6.148. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $r \in \{1,2,\ldots,m\}$. Let $B \in \mathbb{K}^{n\times m}$. Then, $(B^T)_{\sim r,\bullet} = (B_{\bullet,\sim r})^T$.

Proof of Lemma 6.148. Taking the transpose of a matrix turns all its columns into rows. Thus, if we remove the r-th **column** from B and then take the transpose of the resulting matrix, then we obtain the same matrix as if we first take the transpose of B and then remove the r-th **row** from it. Translating this statement into formulas, we obtain precisely $(B_{\bullet,\sim r})^T = (B^T)_{\sim r,\bullet}$. Thus, Lemma 6.148 is proven.[278] □

[277] Recall that B^T denotes the transpose of the matrix B (see Definition 6.10).

[278] A more formal proof of this could be given using Proposition 6.79 **(e)**.

Proof of Proposition 6.146. **(b)** Let $q \in \{1, 2, \ldots, n\}$. Thus, $q \in \{1, 2, \ldots, n\} = \{1, 2, \ldots, (n+1) - 1\}$ (since $n = (n+1) - 1$).

We have $B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n+1}$. Thus, the definition of B^T yields

$$B^T = \left(b_{j,i}\right)_{1 \leq i \leq n+1,\ 1 \leq j \leq n} = \left(b_{j,i}\right)_{1 \leq i \leq n+1,\ 1 \leq j \leq (n+1)-1}$$

(since $n = (n+1) - 1$). Also, $B^T = \left(b_{j,i}\right)_{1 \leq i \leq n+1,\ 1 \leq j \leq (n+1)-1} \in \mathbb{K}^{(n+1) \times ((n+1)-1)}$. Thus, Proposition 6.144 **(b)** (applied to $n+1$, B^T and $b_{j,i}$ instead of n, B and $b_{i,j}$) yields

$$\sum_{r=1}^{n+1} (-1)^r \det\left(\left(B^T\right)_{\sim r, \bullet}\right) b_{q,r} = 0. \qquad (468)$$

But every $r \in \{1, 2, \ldots, n+1\}$ satisfies

$$\det\left(\left(B^T\right)_{\sim r, \bullet}\right) = \det\left(B_{\bullet, \sim r}\right) \qquad (469)$$

[279]. Hence, (468) yields

$$0 = \sum_{r=1}^{n+1} (-1)^r \underbrace{\det\left(\left(B^T\right)_{\sim r, \bullet}\right)}_{\substack{=\det(B_{\bullet,\sim r}) \\ \text{(by (469))}}} b_{q,r} = \sum_{r=1}^{n+1} (-1)^r \det\left(B_{\bullet, \sim r}\right) b_{q,r}.$$

This proves Proposition 6.146 **(b)**.

(a) Write the matrix B in the form $B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n+1}$. For every $r \in \{1, 2, \ldots, n+1\}$, we have

$$\begin{aligned} B_{\bullet, r} &= (\text{the } r\text{-th column of the matrix } B) \\ &\qquad \left(\begin{array}{c}\text{since } B_{\bullet,r} \text{ is the } r\text{-th column of the matrix } B \\ \text{(by the definition of } B_{\bullet,r}\text{)}\end{array}\right) \\ &= \begin{pmatrix} b_{1,r} \\ b_{2,r} \\ \vdots \\ b_{n,r} \end{pmatrix} \qquad \left(\text{since } B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n+1}\right) \\ &= \left(b_{i,r}\right)_{1 \leq i \leq n,\ 1 \leq j \leq 1}. \end{aligned}$$

[279] *Proof of (469):* Let $r \in \{1, 2, \ldots, n+1\}$. Then, $B_{\bullet, \sim r} \in \mathbb{K}^{n \times n}$ (since $B \in \mathbb{K}^{n \times (n+1)}$). In other words, $B_{\bullet, \sim r}$ is an $n \times n$-matrix. Thus, Exercise 6.4 (applied to $A = B_{\bullet, \sim r}$) yields $\det\left(\left(B_{\bullet, \sim r}\right)^T\right) = \det\left(B_{\bullet, \sim r}\right)$.

But Lemma 6.148 (applied to $m = n+1$) yields $\left(B^T\right)_{\sim r, \bullet} = \left(B_{\bullet, \sim r}\right)^T$. Thus, $\det\left(\underbrace{\left(B^T\right)_{\sim r, \bullet}}_{=\left(B_{\bullet, \sim r}\right)^T}\right) = \det\left(\left(B_{\bullet, \sim r}\right)^T\right) = \det\left(B_{\bullet, \sim r}\right)$. This proves (469).

Thus,

$$\begin{aligned}
&\sum_{r=1}^{n+1} (-1)^r \det\left(B_{\bullet,\sim r}\right) \underbrace{B_{\bullet,r}}_{=\left(b_{i,r}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}} \\
&= \sum_{r=1}^{n+1} \underbrace{(-1)^r \det\left(B_{\bullet,\sim r}\right) \left(b_{i,r}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}}_{=\left((-1)^r \det(B_{\bullet,\sim r}) b_{i,r}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}} = \sum_{r=1}^{n+1} \left((-1)^r \det\left(B_{\bullet,\sim r}\right) b_{i,r}\right)_{1\leq i\leq n,\ 1\leq j\leq 1} \\
&= \left(\underbrace{\sum_{r=1}^{n+1} (-1)^r \det\left(B_{\bullet,\sim r}\right) b_{i,r}}_{\substack{=0 \\ \text{(by Proposition 6.146 (b), applied to } q=i\text{)}}} \right)_{1\leq i\leq n,\ 1\leq j\leq 1} = (0)_{1\leq i\leq n,\ 1\leq j\leq 1} = 0_{n\times 1}.
\end{aligned}$$

This proves Proposition 6.146 **(a)**. □

Remark 6.149. Proposition 6.146 **(a)** can be viewed as a restatement of Cramer's rule (Theorem 6.120 **(a)**). More precisely, it is easy to derive one of these two facts from the other (although neither of the two is difficult to prove to begin with). Let us sketch one direction of this argument: namely, let us derive Theorem 6.120 **(a)** from Proposition 6.146 **(a)**.

Indeed, let n, A, $b = (b_1, b_2, \ldots, b_n)^T$ and $A_j^\#$ be as in Theorem 6.120 **(a)**. We want to prove that $A \cdot \left(\det\left(A_1^\#\right), \det\left(A_2^\#\right), \ldots, \det\left(A_n^\#\right)\right)^T = \det A \cdot b$.

Let $B = (A \mid b)$ (using the notations of Definition 6.132); this is an $n \times (n+1)$-matrix.

Fix $r \in \{1, 2, \ldots, n\}$. The matrix $B_{\bullet,\sim r}$ differs from the matrix $A_r^\#$ only in the order of its columns: More precisely,

- the matrix $B_{\bullet,\sim r}$ is obtained from the matrix A by removing the r-th column and attaching the column vector b to the right edge, whereas
- the matrix $A_r^\#$ is obtained from the matrix A by replacing the r-th column by the column vector b.

Thus, the matrix $B_{\bullet,\sim r}$ can be obtained from the matrix $A_r^\#$ by first switching the r-th and $(r+1)$-th columns, then switching the $(r+1)$-th and $(r+2)$-th columns, etc., until finally switching the $(n-1)$-th and n-th columns. Each of these switches multiplies the determinant by -1 (by Exercise 6.7 **(b)**); thus, our sequence of switches multiplies the determinant by $(-1)^{n-r} = (-1)^{n+r}$. Hence,

$$\det\left(B_{\bullet,\sim r}\right) = (-1)^{n+r} \det\left(A_r^\#\right). \tag{470}$$

Now, forget that we fixed r. It is easy to see that every $(v_1, v_2, \ldots, v_n)^T \in \mathbb{K}^{1\times n}$ satisfies

$$A \cdot (v_1, v_2, \ldots, v_n)^T = \sum_{r=1}^{n} v_r A_{\bullet,r}.$$

Applying this to $(v_1, v_2, \ldots, v_n)^T = \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T$, we obtain

$$A \cdot \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T = \sum_{r=1}^{n} \det\left(A_r^{\#}\right) A_{\bullet,r}. \qquad (471)$$

But Proposition 6.146 **(a)** yields

$$\begin{aligned}
0_{n\times 1} &= \sum_{r=1}^{n+1} (-1)^r \det\left(B_{\bullet,\sim r}\right) B_{\bullet,r} \\
&= \sum_{r=1}^{n} (-1)^r \underbrace{\det\left(B_{\bullet,\sim r}\right)}_{\substack{=(-1)^{n+r}\det\left(A_r^{\#}\right) \\ \text{(by (470))}}} \underbrace{B_{\bullet,r}}_{\substack{=A_{\bullet,r} \\ \text{(since } B=(A|b) \text{ and } r\leq n)}} \\
&\qquad + (-1)^{n+1} \det\left(\underbrace{B_{\bullet,\sim(n+1)}}_{\substack{=A \\ \text{(since } B=(A|b))}}\right) \underbrace{B_{\bullet,n+1}}_{\substack{=b \\ \text{(since } B=(A|b))}} \\
&= \sum_{r=1}^{n} \underbrace{(-1)^r (-1)^{n+r}}_{=(-1)^n} \det\left(A_r^{\#}\right) A_{\bullet,r} + \underbrace{(-1)^{n+1}}_{=-(-1)^n} \det A \cdot b \\
&= \sum_{r=1}^{n} (-1)^n \det\left(A_r^{\#}\right) A_{\bullet,r} - (-1)^n \det A \cdot b \\
&= (-1)^n \left(\sum_{r=1}^{n} \det\left(A_r^{\#}\right) A_{\bullet,r} - \det A \cdot b\right).
\end{aligned}$$

Multiplying both sides of this equality by $(-1)^n$, we obtain

$$\begin{aligned}
0_{n\times 1} &= \underbrace{(-1)^n (-1)^n}_{=1} \left(\sum_{r=1}^{n} \det\left(A_r^{\#}\right) A_{\bullet,r} - \det A \cdot b\right) \\
&= \sum_{r=1}^{n} \det\left(A_r^{\#}\right) A_{\bullet,r} - \det A \cdot b.
\end{aligned}$$

Hence,

$$\det A \cdot b = \sum_{r=1}^{n} \det\left(A_r^{\#}\right) A_{\bullet,r} = A \cdot \left(\det\left(A_1^{\#}\right), \det\left(A_2^{\#}\right), \ldots, \det\left(A_n^{\#}\right)\right)^T$$

(by (471)). Thus, we have derived Theorem 6.120 **(a)** from Proposition 6.146 **(a)**. Essentially the same argument (but read backwards) can be used to derive Proposition 6.146 **(a)** from Theorem 6.120 **(a)**.

Now, we can easily prove the *Plücker identity*:

Theorem 6.150. Let n be a positive integer. Let $A \in \mathbb{K}^{n\times(n-1)}$ and $B \in \mathbb{K}^{n\times(n+1)}$. Then,

$$\sum_{r=1}^{n+1} (-1)^r \det(A \mid B_{\bullet,r}) \det(B_{\bullet,\sim r}) = 0$$

(where we are using the notations from Definition 6.132 and from Definition 6.128).

Example 6.151. If $n = 3$, $A = \begin{pmatrix} a & a' \\ b & b' \\ c & c' \end{pmatrix}$ and $B = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{pmatrix}$, then Theorem 6.150 says that

$$\begin{aligned} & -\det\begin{pmatrix} a & a' & x_1 \\ b & b' & y_1 \\ c & c' & z_1 \end{pmatrix} \det\begin{pmatrix} x_2 & x_3 & x_4 \\ y_2 & y_3 & y_4 \\ z_2 & z_3 & z_4 \end{pmatrix} \\ & \qquad + \det\begin{pmatrix} a & a' & x_2 \\ b & b' & y_2 \\ c & c' & z_2 \end{pmatrix} \det\begin{pmatrix} x_1 & x_3 & x_4 \\ y_1 & y_3 & y_4 \\ z_1 & z_3 & z_4 \end{pmatrix} \\ & \qquad - \det\begin{pmatrix} a & a' & x_3 \\ b & b' & y_3 \\ c & c' & z_3 \end{pmatrix} \det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ z_1 & z_2 & z_4 \end{pmatrix} \\ & \qquad + \det\begin{pmatrix} a & a' & x_4 \\ b & b' & y_4 \\ c & c' & z_4 \end{pmatrix} \det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} \\ & = 0. \end{aligned}$$

Proof of Theorem 6.150. Write the matrix B in the form $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n+1}$.

Let $r \in \{1, 2, \ldots, n+1\}$. Then,

$$\begin{aligned}
B_{\bullet,r} &= (\text{the } r\text{-th column of the matrix } B) \\
&\qquad \left(\begin{array}{c}\text{since } B_{\bullet,r} \text{ is the } r\text{-th column of the matrix } B \\ \text{(by the definition of } B_{\bullet,r}\text{)}\end{array}\right) \\
&= \begin{pmatrix} b_{1,r} \\ b_{2,r} \\ \vdots \\ b_{n,r} \end{pmatrix} \qquad \left(\text{since } B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n+1}\right) \\
&= \left(b_{1,r}, b_{2,r}, \ldots, b_{n,r}\right)^T .
\end{aligned}$$

Hence, Proposition 6.135 **(a)** (applied to $B_{\bullet,r}$ and $b_{i,r}$ instead of v and v_i) shows that

$$\det\left(A \mid B_{\bullet,r}\right) = \sum_{i=1}^{n} (-1)^{n+i} b_{i,r} \det\left(A_{\sim i,\bullet}\right). \tag{472}$$

Now, forget that we fixed r. We thus have proven (472) for each $r \in \{1, 2, \ldots, n+1\}$. Now,

$$\begin{aligned}
&\sum_{r=1}^{n+1} (-1)^r \underbrace{\det\left(A \mid B_{\bullet,r}\right)}_{\substack{=\sum_{i=1}^{n}(-1)^{n+i}b_{i,r}\det\left(A_{\sim i,\bullet}\right) \\ \text{(by (472))}}} \det\left(B_{\bullet,\sim r}\right) \\
&= \sum_{r=1}^{n+1} (-1)^r \left(\sum_{i=1}^{n} (-1)^{n+i} b_{i,r} \det\left(A_{\sim i,\bullet}\right)\right) \det\left(B_{\bullet,\sim r}\right) \\
&= \underbrace{\sum_{r=1}^{n+1} \sum_{i=1}^{n}}_{=\sum_{i=1}^{n}\sum_{r=1}^{n+1}} \underbrace{(-1)^r (-1)^{n+i} b_{i,r} \det\left(A_{\sim i,\bullet}\right) \det\left(B_{\bullet,\sim r}\right)}_{=(-1)^r \det(B_{\bullet,\sim r}) b_{i,r}\cdot(-1)^{n+i}\det\left(A_{\sim i,\bullet}\right)} \\
&= \sum_{i=1}^{n} \sum_{r=1}^{n+1} (-1)^r \det\left(B_{\bullet,\sim r}\right) b_{i,r} \cdot (-1)^{n+i} \det\left(A_{\sim i,\bullet}\right) \\
&= \sum_{i=1}^{n} \underbrace{\left(\sum_{r=1}^{n+1} (-1)^r \det\left(B_{\bullet,\sim r}\right) b_{i,r}\right)}_{\substack{=0 \\ \text{(by Proposition 6.146 \textbf{(b)},} \\ \text{applied to } q=i\text{)}}} (-1)^{n+i} \det\left(A_{\sim i,\bullet}\right) \\
&= \sum_{i=1}^{n} 0 (-1)^{n+i} \det\left(A_{\sim i,\bullet}\right) = 0.
\end{aligned}$$

This proves Theorem 6.150. □

Remark 6.152. Theorem 6.150 (at least in the case when $\mathbb{K}$ is a field) is essentially equivalent to [Fulton97, §9.1, Exercise 1], to [KleLak72, (QR)], to [Jacobs10, Theorem 3.4.11 (the "necessary" part)], and to [Lampe13, Proposition 3.3.2 (the "only if" part)].

Exercise 6.43. Use Theorem 6.150 to give a new proof of Proposition 6.137.

6.22. Laplace expansion in multiple rows/columns

In this section, we shall see a (somewhat unwieldy, but classical and important) generalization of Theorem 6.82. First, we shall need some notations:

Definition 6.153. Throughout Section 6.22, we shall use the following notations:

- If I is a finite set of integers, then $\sum I$ shall denote the sum of all elements of I. (Thus, $\sum I = \sum_{i \in I} i$.)
- If I is a finite set of integers, then $w(I)$ shall denote the list of all elements of I in increasing order (with no repetitions). (See Definition 2.50 for the formal definition of this list.) (For example, $w(\{3,4,8\}) = (3,4,8)$.)

We shall also use the notation introduced in Definition 6.78. If n, m, A, $(i_1, i_2, \ldots, i_u)$ and $(j_1, j_2, \ldots, j_v)$ are as in Definition 6.78, then we shall use the notation $\operatorname{sub}_{(i_1,i_2,\ldots,i_u)}^{(j_1,j_2,\ldots,j_v)} A$ as a synonym for $\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v} A$.

A consequence of this definition is that if A is an $n \times m$-matrix, and if U is a subset of $\{1,2,\ldots,n\}$, and if V is a subset of $\{1,2,\ldots,m\}$, then $\operatorname{sub}_{w(U)}^{w(V)} A$ is a well-defined $|U| \times |V|$-matrix[280] (actually, a submatrix of A).

Example 6.154. If $n = 3$ and $m = 4$ and $A = \begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{pmatrix}$, then $\operatorname{sub}_{w(\{2,3\})}^{w(\{1,3,4\})} A = \begin{pmatrix} a' & c' & d' \\ a'' & c'' & d'' \end{pmatrix}$.

The following fact is obvious from the definition of $w(I)$:

Proposition 6.155. Let I be a finite set of integers. Then, $w(I)$ is an $|I|$-tuple of elements of I.

[280] because $w(U)$ is a list of $|U|$ elements of $\{1,2,\ldots,n\}$, and because $w(V)$ is a list of $|V|$ elements of $\{1,2,\ldots,m\}$

Theorem 6.156. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times n}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. (For instance, if $n = 4$ and $I = \{1, 4\}$, then $\widetilde{I} = \{2, 3\}$.)

(a) For every subset P of $\{1, 2, \ldots, n\}$, we have

$$\det A = \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |Q| = |P|}} (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A \right).$$

(b) For every subset Q of $\{1, 2, \ldots, n\}$, we have

$$\det A = \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A \right).$$

Theorem 6.156 is actually a generalization of Theorem 6.82, known as "Laplace expansion in multiple rows (resp. columns)". It appears (for example) in [Willia18, Theorem 3.61] and in [Prasol94, Theorem 2.4.1].[281] Theorem 6.82 **(a)** can be recovered from Theorem 6.156 **(a)** by setting $P = \{p\}$; similarly for the part **(b)**.

Example 6.157. Let us see what Theorem 6.156 **(a)** says in a simple case. For this example, set $n = 4$ and $A = \left(a_{i,j}\right)_{1 \leq i \leq 4,\ 1 \leq j \leq 4}$. Also, set $P = \{1, 4\} \subseteq \{1, 2, 3, 4\}$; thus, $w(P) = (1, 4)$, $\sum P = 1 + 4 = 5$, $|P| = 2$, $\widetilde{P} = \{2, 3\}$ and $w\left(\widetilde{P}\right) = (2, 3)$.

[281] Of course, parts **(a)** and **(b)** of Theorem 6.156 are easily seen to be equivalent; thus, many authors confine themselves to only stating one of them. For example, Theorem 6.156 is [CaSoSp12, Lemma A.1 (f)].

Now, Theorem 6.156 **(a)** says that

$$\begin{aligned}
&\det A \\
&= \sum_{\substack{Q\subseteq\{1,2,3,4\};\\ |Q|=2}} (-1)^{5+\sum Q} \det\left(\operatorname{sub}_{1,4}^{w(Q)} A\right) \det\left(\operatorname{sub}_{2,3}^{w(\widetilde{Q})} A\right) \\
&= (-1)^{5+(1+2)} \det\left(\operatorname{sub}_{1,4}^{1,2} A\right) \det\left(\operatorname{sub}_{2,3}^{3,4} A\right) \\
&\qquad + (-1)^{5+(1+3)} \det\left(\operatorname{sub}_{1,4}^{1,3} A\right) \det\left(\operatorname{sub}_{2,3}^{2,4} A\right) \\
&\qquad + (-1)^{5+(1+4)} \det\left(\operatorname{sub}_{1,4}^{1,4} A\right) \det\left(\operatorname{sub}_{2,3}^{2,3} A\right) \\
&\qquad + (-1)^{5+(2+3)} \det\left(\operatorname{sub}_{1,4}^{2,3} A\right) \det\left(\operatorname{sub}_{2,3}^{1,4} A\right) \\
&\qquad + (-1)^{5+(2+4)} \det\left(\operatorname{sub}_{1,4}^{2,4} A\right) \det\left(\operatorname{sub}_{2,3}^{1,3} A\right) \\
&\qquad + (-1)^{5+(3+4)} \det\left(\operatorname{sub}_{1,4}^{3,4} A\right) \det\left(\operatorname{sub}_{2,3}^{1,2} A\right) \\
&= \det\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{4,1} & a_{4,2} \end{pmatrix} \det\begin{pmatrix} a_{2,3} & a_{2,4} \\ a_{3,3} & a_{3,4} \end{pmatrix} \\
&\qquad - \det\begin{pmatrix} a_{1,1} & a_{1,3} \\ a_{4,1} & a_{4,3} \end{pmatrix} \det\begin{pmatrix} a_{2,2} & a_{2,4} \\ a_{3,2} & a_{3,4} \end{pmatrix} \\
&\qquad + \det\begin{pmatrix} a_{1,1} & a_{1,4} \\ a_{4,1} & a_{4,4} \end{pmatrix} \det\begin{pmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} \\
&\qquad + \det\begin{pmatrix} a_{1,2} & a_{1,3} \\ a_{4,2} & a_{4,3} \end{pmatrix} \det\begin{pmatrix} a_{2,1} & a_{2,4} \\ a_{3,1} & a_{3,4} \end{pmatrix} \\
&\qquad - \det\begin{pmatrix} a_{1,2} & a_{1,4} \\ a_{4,2} & a_{4,4} \end{pmatrix} \det\begin{pmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{pmatrix} \\
&\qquad + \det\begin{pmatrix} a_{1,3} & a_{1,4} \\ a_{4,3} & a_{4,4} \end{pmatrix} \det\begin{pmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix}.
\end{aligned}$$

The following lemma will play a crucial role in our proof of Theorem 6.156 (similar to the role that Lemma 6.84 played in our proof of Theorem 6.82):

Lemma 6.158. Let $n \in \mathbb{N}$. For any subset I of $\{1,2,\ldots,n\}$, we let $\widetilde{I}$ denote the complement $\{1,2,\ldots,n\} \setminus I$ of I.

Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be two $n\times n$-matrices. Let

P and Q be two subsets of $\{1, 2, \ldots, n\}$ such that $|P| = |Q|$. Then,

$$\sum_{\substack{\sigma \in S_n; \\ \sigma(P) = Q}} (-1)^{\sigma} \left(\prod_{i \in P} a_{i,\sigma(i)} \right) \left(\prod_{i \in \widetilde{P}} b_{i,\sigma(i)} \right)$$
$$= (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B \right).$$

The proof of Lemma 6.158 is similar (in its spirit) to the proof of Lemma 6.84, but it requires a lot more bookkeeping (if one wants to make it rigorous). This proof shall be given in the solution to Exercise 6.44:

Exercise 6.44. Prove Lemma 6.158 and Theorem 6.156.

[Hint: First, prove Lemma 6.158 in the case when $P = \{1, 2, \ldots, k\}$ for some $k \in \{0, 1, \ldots, n\}$; in order to do so, use the bijection from Exercise 5.14 **(c)** (applied to $I = Q$). Then, derive the general case of Lemma 6.158 by permuting the rows of the matrices. Finally, prove Theorem 6.156.]

The following exercise generalizes Proposition 6.96 in the same way as Theorem 6.156 generalizes Theorem 6.82:

Exercise 6.45. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times n}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. Let R be a subset of $\{1, 2, \ldots, n\}$. Prove the following:

(a) For every subset P of $\{1, 2, \ldots, n\}$ satisfying $|P| = |R|$ and $P \neq R$, we have

$$0 = \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |Q| = |P|}} (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(R)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A \right).$$

(b) For every subset Q of $\{1, 2, \ldots, n\}$ satisfying $|Q| = |R|$ and $Q \neq R$, we have

$$0 = \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(R)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A \right).$$

Exercise 6.45 can be generalized to non-square matrices:

Exercise 6.46. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. Let J and K be two subsets of $\{1, 2, \ldots, m\}$ satisfying $|J| + |K| = n$ and $J \cap K \neq \varnothing$.

(a) For every $A \in \mathbb{K}^{m \times n}$, we have

$$0 = \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |Q| = |J|}} (-1)^{\sum Q} \det \left(\operatorname{sub}_{w(J)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(K)}^{w(\widetilde{Q})} A \right).$$

(b) For every $A \in \mathbb{K}^{n \times m}$, we have

$$0 = \sum_{\substack{P \subseteq \{1,2,\ldots,n\};\\ |P|=|J|}} (-1)^{\sum P} \det\left(\operatorname{sub}_{w(P)}^{w(J)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(K)} A\right).$$

(Exercise 6.45 can be obtained as a particular case of Exercise 6.46; we leave the details to the reader.)

The following exercise gives a first application of Theorem 6.156 (though it can also be solved with more elementary methods):

Exercise 6.47. Let $n \in \mathbb{N}$. Let P and Q be two subsets of $\{1, 2, \ldots, n\}$. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n} \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix such that

$$\text{every } i \in P \text{ and } j \in Q \text{ satisfy } a_{i,j} = 0. \tag{473}$$

Prove the following:

(a) If $|P| + |Q| > n$, then $\det A = 0$.

(b) If $|P| + |Q| = n$, then

$$\det A = (-1)^{\sum P + \sum \widetilde{Q}} \det\left(\operatorname{sub}_{w(P)}^{w(\widetilde{Q})} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(Q)} A\right).$$

Example 6.159. **(a)** Applying Exercise 6.47 **(a)** to $n = 5$, $P = \{1, 3, 5\}$ and $Q = \{2, 3, 4\}$, we see that

$$\det \begin{pmatrix} a & 0 & 0 & 0 & b \\ c & d & e & f & g \\ h & 0 & 0 & 0 & i \\ j & k & \ell & m & n \\ o & 0 & 0 & 0 & p \end{pmatrix} = 0$$

(don't mistake the letter "o" for a zero) for any $a, b, c, d, e, f, g, h, i, j, k, \ell, m, n, o, p \in \mathbb{K}$ (since the $n \times n$-matrices $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ satisfying (473) are precisely the matrices of the form $\begin{pmatrix} a & 0 & 0 & 0 & b \\ c & d & e & f & g \\ h & 0 & 0 & 0 & i \\ j & k & \ell & m & n \\ o & 0 & 0 & 0 & p \end{pmatrix}$).

(b) Applying Exercise 6.47 **(a)** to $n = 5$, $P = \{2, 3, 4\}$ and $Q = \{2, 3, 4\}$, we see that

$$\det \begin{pmatrix} a & b & c & d & e \\ f & 0 & 0 & 0 & g \\ h & 0 & 0 & 0 & i \\ j & 0 & 0 & 0 & k \\ \ell & m & n & o & p \end{pmatrix} = 0$$

(don't mistake the letter "o" for a zero) for any $a, b, c, d, e, f, g, h, i, j, k, \ell, m, n, o, p \in \mathbb{K}$. This is precisely the claim of Exercise 6.6 **(b)**.

(c) Applying Exercise 6.47 **(b)** to $n = 4$, $P = \{2, 3\}$ and $Q = \{2, 3\}$, we see that

$$\det \begin{pmatrix} a & b & c & d \\ \ell & 0 & 0 & e \\ k & 0 & 0 & f \\ j & i & h & g \end{pmatrix} = \det \begin{pmatrix} \ell & e \\ k & f \end{pmatrix} \cdot \det \begin{pmatrix} b & c \\ i & h \end{pmatrix}$$

for all $a, b, c, d, e, f, g, h, i, j, k, \ell \in \mathbb{K}$. This solves Exercise 6.6 **(a)**.

(d) It is not hard to derive Exercise 6.29 by applying Exercise 6.47 **(b)** to $n + m$, $\begin{pmatrix} A & 0_{n \times m} \\ C & D \end{pmatrix}$, $\{1, 2, \ldots, n\}$ and $\{n + 1, n + 2, \ldots, n + m\}$ instead of n, A, P and Q. Similarly, Exercise 6.30 can be derived from Exercise 6.47 **(b)** as well.

6.23. $\det(A + B)$

As Theorem 6.23 shows, the determinant of the product AB of two square matrices can be easily and neatly expressed through the determinants of A and B. In contrast, the determinant of a sum $A + B$ of two square matrices cannot be expressed in such a way[282]. There is, however, a formula for $\det(A + B)$ in terms of the determinants of submatrices of A and B. While it is rather unwieldy (a far cry from the elegance of Theorem 6.23), it is nevertheless useful sometimes; let us now show it:

Theorem 6.160. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. (For instance, if $n = 4$ and $I = \{1, 4\}$, then $\widetilde{I} = \{2, 3\}$.) Let us use the notations introduced in Definition 6.78 and in Definition 6.153.

Let A and B be two $n \times n$-matrices. Then,

$$\det(A + B) = \sum_{P \subseteq \{1,2,\ldots,n\}} \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} (-1)^{\sum P + \sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right).$$

Example 6.161. For this example, set $n = 2$, $A = (a_{i,j})_{1 \leq i \leq 2,\ 1 \leq j \leq 2}$ and $B =$

[282] It is easy to find two 2×2-matrices A_1 and B_1 and two other 2×2-matrices A_2 and B_2 such that $\det(A_1) = \det(A_2)$ and $\det(B_1) = \det(B_2)$ but $\det(A_1 + B_1) \neq \det(A_2 + B_2)$. This shows that $\det(A + B)$ cannot generally be computed from $\det A$ and $\det B$.

$\left(b_{i,j}\right)_{1\leq i\leq 2,\ 1\leq j\leq 2}$. Then, Theorem 6.160 says that

$$\begin{aligned}
&\det(A+B) \\
&= \sum_{P\subseteq\{1,2\}} \sum_{\substack{Q\subseteq\{1,2\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right) \\
&= (-1)^{\sum\varnothing+\sum\varnothing} \det \underbrace{(\operatorname{sub} A)}_{\substack{\text{this is the}\\ 0\times 0\text{-matrix}}} \det\left(\operatorname{sub}_{1,2}^{1,2} B\right) \\
&\qquad + (-1)^{\sum\{1\}+\sum\{1\}} \det\left(\operatorname{sub}_1^1 A\right) \det\left(\operatorname{sub}_2^2 B\right) \\
&\qquad + (-1)^{\sum\{1\}+\sum\{2\}} \det\left(\operatorname{sub}_1^2 A\right) \det\left(\operatorname{sub}_2^1 B\right) \\
&\qquad + (-1)^{\sum\{2\}+\sum\{1\}} \det\left(\operatorname{sub}_2^1 A\right) \det\left(\operatorname{sub}_1^2 B\right) \\
&\qquad + (-1)^{\sum\{2\}+\sum\{2\}} \det\left(\operatorname{sub}_2^2 A\right) \det\left(\operatorname{sub}_1^1 B\right) \\
&\qquad + (-1)^{\sum\{1,2\}+\sum\{1,2\}} \det\left(\operatorname{sub}_{1,2}^{1,2} A\right) \det \underbrace{(\operatorname{sub} B)}_{\substack{\text{this is the}\\ 0\times 0\text{-matrix}}} \\
&= \underbrace{\det(\text{the } 0\times 0\text{-matrix})}_{=1} \det\begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix} + \det\begin{pmatrix} a_{1,1} \end{pmatrix} \det\begin{pmatrix} b_{2,2} \end{pmatrix} \\
&\qquad - \det\begin{pmatrix} a_{1,2} \end{pmatrix} \det\begin{pmatrix} b_{2,1} \end{pmatrix} - \det\begin{pmatrix} a_{2,1} \end{pmatrix} \det\begin{pmatrix} b_{1,2} \end{pmatrix} \\
&\qquad + \det\begin{pmatrix} a_{2,2} \end{pmatrix} \det\begin{pmatrix} b_{1,1} \end{pmatrix} + \det\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \underbrace{\det(\text{the } 0\times 0\text{-matrix})}_{=1} \\
&= \det\begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix} + a_{1,1}b_{2,2} - a_{1,2}b_{2,1} - a_{2,1}b_{1,2} + a_{2,2}b_{1,1} + \det\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}.
\end{aligned}$$

Exercise 6.48. Prove Theorem 6.160.
[**Hint:** Use Lemma 6.158.]

Theorem 6.160 takes a simpler form in the particular case when the matrix B is diagonal (i.e., has all entries outside of its diagonal equal to 0):

Corollary 6.162. Let $n \in \mathbb{N}$. For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

Let A be an $n\times n$-matrix. Let $d_1, d_2, \ldots, d_n$ be n elements of $\mathbb{K}$. Let D be the $n\times n$-matrix $\left(d_i\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Then,

$$\det(A+D) = \sum_{P\subseteq\{1,2,\ldots,n\}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i.$$

This corollary can easily be derived from Theorem 6.160 using the following fact:

Lemma 6.163. Let $n \in \mathbb{N}$. For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

Let $d_1, d_2, \ldots, d_n$ be n elements of $\mathbb{K}$. Let D be the $n \times n$-matrix $\left(d_i \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Let P and Q be two subsets of $\{1, 2, \ldots, n\}$ such that $|P| = |Q|$. Then,

$$\det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = \delta_{P,Q} \prod_{i \in P} d_i.$$

Proving Corollary 6.162 and Lemma 6.163 in detail is part of Exercise 6.49 further below.

A particular case of Corollary 6.162 is the following fact:

Corollary 6.164. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let $x \in \mathbb{K}$. Then,

$$\det\left(A + xI_n\right) = \sum_{P \subseteq \{1,2,\ldots,n\}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) x^{n-|P|} \tag{474}$$

$$= \sum_{k=0}^{n} \left(\sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \right) x^k. \tag{475}$$

Exercise 6.49. Prove Corollary 6.162, Lemma 6.163 and Corollary 6.164.

Remark 6.165. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix over the commutative ring $\mathbb{K}$. Consider the commutative ring $\mathbb{K}[X]$ of polynomials in the indeterminate X over $\mathbb{K}$ (that is, polynomials in the indeterminate X with coefficients lying in $\mathbb{K}$). We can then regard A as a matrix over the ring $\mathbb{K}[X]$ as well (because every element of $\mathbb{K}$ can be viewed as a constant polynomial in $\mathbb{K}[X]$).

Consider the $n \times n$-matrix $A + XI_n$ over the commutative ring $\mathbb{K}[X]$. (For example, if $n = 2$ and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $A + XI_n = \begin{pmatrix} a & b \\ c & d \end{pmatrix} + X \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a+X & b \\ c & d+X \end{pmatrix}$. In general, the matrix $A + XI_n$ is obtained from A by adding an X to each diagonal entry.)

The determinant $\det\left(A + XI_n\right)$ is a polynomial in $\mathbb{K}[X]$. (For instance, for

$n = 2$ and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we have

$$\det(A + XI_n) = \det \begin{pmatrix} a+X & b \\ c & d+X \end{pmatrix} = (a+X)(d+X) - bc$$
$$= X^2 + (a+d)X + (ad - bc).$$

)

This polynomial $\det(A + XI_n)$ is a highly important object; it is a close relative of what is called the *characteristic polynomial* of A. (More precisely, the characteristic polynomial of A is either $\det(XI_n - A)$ or $\det(A - XI_n)$, depending on the conventions that one is using; thus, the polynomial $\det(A + XI_n)$ is either the characteristic polynomial of $-A$ or $(-1)^n$ times this characteristic polynomial.) For more about the characteristic polynomial, see [Artin10, Section 4.5] or [Heffer17, Chapter Five, Section II, §3] (or various other texts on linear algebra).

Using Corollary 6.164, we can explicitly compute the coefficients of the polynomial $\det(A + XI_n)$. In fact, (475) (applied to $\mathbb{K}[X]$ and X instead of $\mathbb{K}$ and x) yields

$$\det(A + XI_n) = \sum_{k=0}^{n} \left(\sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \right) X^k.$$

Hence, for every $k \in \{0, 1, \ldots, n\}$, the coefficient of X^k in the polynomial $\det(A + XI_n)$ is

$$\sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right).$$

In particular:

- The coefficient of X^n in the polynomial $\det(A + XI_n)$ is

$$\sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-n}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right)$$
$$= \det \underbrace{\left(\operatorname{sub}_{w(\varnothing)}^{w(\varnothing)} A\right)}_{=(\text{the } 0\times 0\text{-matrix})} \qquad \left(\begin{array}{c}\text{since the only subset } P \text{ of } \{1,2,\ldots,n\} \\ \text{satisfying } |P| = n-n \text{ is the empty set } \varnothing\end{array}\right)$$
$$= \det(\text{the } 0 \times 0\text{-matrix}) = 1.$$

- The coefficient of X^0 in the polynomial $\det(A + XI_n)$ is

$$\sum_{\substack{P \subseteq \{1,2,\ldots,n\};\\ |P| = n-0}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right)$$
$$= \det \underbrace{\left(\operatorname{sub}_{w(\{1,2,\ldots,n\})}^{w(\{1,2,\ldots,n\})} A\right)}_{=\operatorname{sub}_{1,2,\ldots,n}^{1,2,\ldots,n} A = A} \qquad \left(\begin{array}{c}\text{since the only subset } P \text{ of } \{1,2,\ldots,n\} \\ \text{satisfying } |P| = n-0 \text{ is the set } \{1,2,\ldots,n\}\end{array}\right)$$
$$= \det A.$$

- Write the matrix A as $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$. Then, the coefficient of X^{n-1} in the polynomial $\det(A + XI_n)$ is

$$\sum_{\substack{P \subseteq \{1,2,\ldots,n\};\\ |P| = n-(n-1)}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right)$$
$$= \sum_{\substack{P \subseteq \{1,2,\ldots,n\};\\ |P| = 1}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) = \sum_{k=1}^{n} \det \underbrace{\left(\operatorname{sub}_{w(\{k\})}^{w(\{k\})} A\right)}_{\substack{=\operatorname{sub}_k^k A = \left(\begin{array}{c} a_{k,k} \end{array}\right) \\ \text{(this is a } 1\times 1\text{-matrix)}}}$$
$$\left(\begin{array}{c}\text{since the subsets } P \text{ of } \{1,2,\ldots,n\} \\ \text{satisfying } |P| = 1 \text{ are the sets } \{1\},\{2\},\ldots,\{n\}\end{array}\right)$$
$$= \sum_{k=1}^{n} \underbrace{\det\left(\begin{array}{c} a_{k,k} \end{array}\right)}_{=a_{k,k}} = \sum_{k=1}^{n} a_{k,k}.$$

  In other words, this coefficient is the sum of all diagonal entries of A. This sum is called the *trace* of A, and is denoted by $\operatorname{Tr} A$.

6.24. Some alternating-sum formulas

The next few exercises don't all involve determinants; what they have in common is that they contain alternating sums (i.e., sums where the addend contains a power of -1).

Exercise 6.50. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $\mathbb{L}$ be a noncommutative ring. (Keep in mind that our definition of a "noncommutative ring" includes all commutative rings.)

Let $n \in \mathbb{N}$. The summation sign $\sum_{I \subseteq [n]}$ shall mean $\sum_{I \in \mathcal{P}([n])}$, where $\mathcal{P}([n])$ denotes the powerset of $[n]$.

Let $v_1, v_2, \ldots, v_n$ be n elements of $\mathbb{L}$.

(a) Prove that

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m = \sum_{\substack{f:[m] \to [n]; \\ f \text{ is surjective}}} v_{f(1)} v_{f(2)} \cdots v_{f(m)}$$

for each $m \in \mathbb{N}$.

(b) Prove that

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m = 0$$

for each $m \in \{0, 1, \ldots, n-1\}$.

(c) Prove that

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^n = \sum_{\sigma \in S_n} v_{\sigma(1)} v_{\sigma(2)} \cdots v_{\sigma(n)}.$$

(d) Now, assume that $\mathbb{L}$ is a **commutative** ring. Prove that

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^n = n! \cdot v_1 v_2 \cdots v_n.$$

[**Hint:** First, generalize Lemma 6.22 to the case of a noncommutative ring $\mathbb{K}$.]

Exercise 6.51. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $\mathbb{L}$ be a noncommutative ring. (Keep in mind that our definition of a "noncommutative ring" includes all commutative rings.)

Let $n \in \mathbb{N}$. The summation sign $\sum_{I \subseteq [n]}$ shall mean $\sum_{I \in \mathcal{P}([n])}$, where $\mathcal{P}([n])$ denotes the powerset of $[n]$.

Let $v_1, v_2, \ldots, v_n$ be n elements of $\mathbb{L}$.

(a) Prove that

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(w + \sum_{i \in I} v_i \right)^m = 0$$

for each $m \in \{0, 1, \ldots, n-1\}$ and $w \in \mathbb{L}$.

(b) Prove that

$$\sum_{I\subseteq[n]} (-1)^{n-|I|} \left(w + \sum_{i\in I} v_i\right)^n = \sum_{\sigma\in S_n} v_{\sigma(1)} v_{\sigma(2)} \cdots v_{\sigma(n)}$$

for every $w \in \mathbb{L}$.

(c) Prove that

$$\sum_{I\subseteq[n]} (-1)^{n-|I|} \left(\sum_{i\in I} v_i - \sum_{i\in[n]\setminus I} v_i\right)^m = 0$$

for each $m \in \{0, 1, \ldots, n-1\}$.

(d) Prove that

$$\sum_{I\subseteq[n]} (-1)^{n-|I|} \left(\sum_{i\in I} v_i - \sum_{i\in[n]\setminus I} v_i\right)^n = 2^n \sum_{\sigma\in S_n} v_{\sigma(1)} v_{\sigma(2)} \cdots v_{\sigma(n)}.$$

Note that parts **(b)** and **(c)** of Exercise 6.50 are special cases of parts **(a)** and **(b)** of Exercise 6.51 (obtained by setting $w = 0$).

The following exercise generalizes Exercise 6.50:

Exercise 6.52. For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Let $\mathbb{L}$ be a noncommutative ring. (Keep in mind that our definition of a "noncommutative ring" includes all commutative rings.)

Let G be a finite set. Let H be a subset of G. Let $n \in \mathbb{N}$.

For each $i \in G$ and $j \in [n]$, let $b_{i,j}$ be an element of $\mathbb{L}$. For each $j \in [n]$ and each subset I of G, we define an element $b_{I,j} \in \mathbb{L}$ by $b_{I,j} = \sum_{i\in I} b_{i,j}$.

(a) Prove that

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} b_{I,1} b_{I,2} \cdots b_{I,n} = (-1)^{|G|} \sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n}.$$

(b) If $n < |G \setminus H|$, then prove that

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} b_{I,1} b_{I,2} \cdots b_{I,n} = 0.$$

(c) If $n = |G|$, then prove that

$$\sum_{I\subseteq G} (-1)^{|G\setminus I|} b_{I,1} b_{I,2} \cdots b_{I,n} = \sum_{\substack{f:[n]\to G\\ \text{is bijective}}} b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n}.$$

Remark 6.166. Let $\mathbb{K}$ be a commutative ring. Let $n \in \mathbb{N}$. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n} \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix. Then, the *permanent* $\operatorname{per} A$ of A is defined to be the element

$$\sum_{\sigma \in S_n} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

of $\mathbb{K}$. The concept of a permanent is thus similar to the concept of a determinant (the only difference is that the factor $(-1)^{\sigma}$ is missing from the definition of the permanent); however, it has far fewer interesting properties. One of the properties that it does have is the so-called *Ryser formula* (see, e.g., [Comtet74, §4.9, [9e]]), which says that

$$\operatorname{per} A = (-1)^n \sum_{I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|} \prod_{j=1}^{n} \sum_{i \in I} a_{i,j}.$$

The reader is invited to check that this formula follows from Exercise 6.52 **(c)** (applied to $\mathbb{L} = \mathbb{K}$, $G = \{1, 2, \ldots, n\}$, $H = \varnothing$ and $b_{i,j} = a_{i,j}$).

Exercise 6.53. Let $n \in \mathbb{N}$. Let G be a finite set such that $n < |G|$. For each $i \in G$, let $A_i \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix. Prove that

$$\sum_{I \subseteq G} (-1)^{|I|} \det \left(\sum_{i \in I} A_i \right) = 0.$$

Exercise 6.54. Let $n \in \mathbb{N}$. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix.

(a) Prove that

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\sum_{i=1}^{n} a_{i,\sigma(i)} \right)^k = 0$$

for each $k \in \{0, 1, \ldots, n-2\}$.

(b) Assume that $n \geq 1$. Prove that

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\sum_{i=1}^{n} a_{i,\sigma(i)} \right)^{n-1} = (n-1)! \cdot \sum_{p=1}^{n} \sum_{q=1}^{n} (-1)^{p+q} \det \left(A_{\sim p, \sim q} \right).$$

(Here, we are using the notation introduced in Definition 6.81.)

Exercise 6.55. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$.

Let $m = \binom{n}{2}$.

(a) Prove that

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\sum_{i=1}^{n} a_i b_{\sigma(i)} \right)^k = 0$$

for each $k \in \{0, 1, \ldots, m-1\}$.

(b) Prove that

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\sum_{i=1}^{n} a_i b_{\sigma(i)} \right)^m = \mathbf{m}\left(0, 1, \ldots, n-1\right) \cdot \prod_{1 \leq i < j \leq n} \left(\left(a_i - a_j\right) \left(b_i - b_j\right) \right).$$

Here, we are using the notation introduced in Exercise 6.2.

(c) Let $k \in \mathbb{N}$. Prove that

$$\begin{aligned}
&\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\sum_{i=1}^{n} a_i b_{\sigma(i)} \right)^k \\
&= \sum_{\substack{(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n; \\ g_1 < g_2 < \cdots < g_n; \\ g_1 + g_2 + \cdots + g_n = k}} \mathbf{m}\left(g_1, g_2, \ldots, g_n\right) \\
&\qquad \cdot \det\left(\left(a_i^{g_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) \cdot \det\left(\left(b_i^{g_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right).
\end{aligned}$$

Here, we are using the notation introduced in Exercise 6.2.

Note that Exercise 6.55 generalizes [AndDos10, Exercise 12.13] and [AndDos12, §12.1, Problem 1].

6.25. Additional exercises

Here are a few more additional exercises, with no importance to the rest of the text (and mostly no solutions given).

Exercise 6.56. Let $n \in \mathbb{N}$. For any subset I of $\{1,2,\ldots,n\}$, we let $\widetilde{I}$ denote the complement $\{1,2,\ldots,n\} \setminus I$ of I. (For instance, if $n = 4$ and $I = \{1,4\}$, then $\widetilde{I} = \{2,3\}$.) Let us use the notations introduced in Definition 6.78 and in Definition 6.153.

Let $A \in \mathbb{K}^{n\times n}$ be an invertible matrix. Let P and Q be two subsets of $\{1,2,\ldots,n\}$ satisfying $|P| = |Q|$. Prove that

$$\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) = (-1)^{\sum P + \sum Q} \det A \cdot \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} \left(A^{-1}\right)\right). \tag{476}$$

[**Hint:** Apply Exercise 6.38 to a matrix obtained from A by permuting the rows and permuting the columns.]

Remark 6.167. The claim of Exercise 6.56 is the so-called *Jacobi complementary minor theorem*. It appears, for example, in [Lalond96, (1)] and in [CaSoSp12, Lemma A.1 (e)], and is used rather often when working with determinants (for example, it is used in [LLPT95, Chapter SYM, proof of Proposition (7.5) (5)] and many times in [CaSoSp12]).

The determinant of a submatrix of a matrix A is called a *minor* of A. Thus, in the equality (476), the determinant $\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)$ on the left hand side is a minor of A, whereas the determinant $\det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} \left(A^{-1}\right)\right)$ on the right hand side is a minor of A^{-1}. Thus, roughly speaking, the equality (476) says that any minor of A equals a certain minor of A^{-1} times $\det A$ times a certain sign.

It is instructive to check the particular case of (476) obtained when both P and Q are sets of cardinality $n-1$ (so that $\widetilde{P}$ and $\widetilde{Q}$ are 1-element sets). This particular case turns out to be the statement of Theorem 6.110 **(b)** in disguise.

Exercise 6.38 is the particular case of Exercise 6.56 obtained when $P = \{1,2,\ldots,k\}$ and $Q = \{1,2,\ldots,k\}$.

Exercise 6.57. Let n and k be positive integers such that $k \leq n$. Let $A \in \mathbb{K}^{n\times(n-k)}$ and $B \in \mathbb{K}^{n\times(n+k)}$.

Let us use the notations from Definition 6.128. For any subset I of $\{1,2,\ldots,n+k\}$, we introduce the following five notations:

- Let $\sum I$ denote the sum of all elements of I. (Thus, $\sum I = \sum_{i\in I} i$.)
- Let $w(I)$ denote the list of all elements of I in increasing order (with no repetitions). (See Definition 2.50 for the formal definition of this list.) (For example, $w(\{3,4,8\}) = (3,4,8)$.)
- Let $(A \mid B_{\bullet,I})$ denote the $n \times (n-k+|I|)$-matrix whose columns are $\underbrace{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,n-k}}_{\text{the columns of } A}, B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ (from left to right), where

$(i_1, i_2, \ldots, i_\ell) = w(I)$.

- Let $B_{\bullet,\sim I}$ denote the $n \times (n + k - |I|)$-matrix whose columns are $B_{\bullet,j_1}, B_{\bullet,j_2}, \ldots, B_{\bullet,j_h}$ (from left to right), where $(j_1, j_2, \ldots, j_h) = w(\{1, 2, \ldots, n+k\} \setminus I)$.
  (Using the notations of Definition 6.31, we can rewrite this definition as $B_{\bullet,\sim I} = \operatorname{cols}_{j_1,j_2,\ldots,j_h} B$, where $(j_1, j_2, \ldots, j_h) = w(\{1, 2, \ldots, n+k\} \setminus I)$.)

Then, prove that

$$\sum_{\substack{I \subseteq \{1,2,\ldots,n+k\};\\ |I|=k}} (-1)^{\sum I + (1+2+\cdots+k)} \det(A \mid B_{\bullet,I}) \det(B_{\bullet,\sim I}) = 0.$$

(Note that this generalizes Theorem 6.150; indeed, the latter theorem is the particular case for $k = 1$.)

Exercise 6.58. Recall that the binomial coefficients satisfy the recurrence relation (234), which (visually) says that every entry of Pascal's triangle is the sum of the two entries left-above it and right-above it.

Let us now define a variation of Pascal's triangle as follows: Define a nonnegative integer $\binom{m}{n}_D$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$ recursively as follows:

- Set $\binom{0}{n}_D = \begin{cases} 1, & \text{if } n = 0; \\ 0, & \text{if } n > 0 \end{cases}$ for every $n \in \mathbb{N}$.
- For every $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$, set $\binom{m}{n}_D = 0$ if either m or n is negative.
- For every positive integer m and every $n \in \mathbb{N}$, set
$$\binom{m}{n}_D = \binom{m-1}{n-1}_D + \binom{m-1}{n}_D + \binom{m-2}{n-1}_D.$$

(Thus, if we lay these $\binom{m}{n}_D$ out in the same way as the binomial coefficients $\binom{m}{n}$ in Pascal's triangle, then every entry is the sum of the three entries left-above it, right-above it, and straight above it.)

The integers $\binom{m}{n}_D$ are known as the Delannoy numbers.

(a) Show that

$$\binom{n+m}{n}_D = \sum_{i=0}^{n} \binom{n}{i} \binom{m+i}{n} = \sum_{i=0}^{n} \binom{n}{i} \binom{m}{i} 2^i$$

for every $n \in \mathbb{N}$ and $m \in \mathbb{N}$. (The second equality sign here is a consequence of Proposition 3.39 **(e)**.)

(b) Let $n \in \mathbb{N}$. Let A be the $n \times n$-matrix $\left(\dbinom{i+j-2}{i-1}_D \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (an analogue of the matrix A from Exercise 6.11). Show that

$$\det A = 2^{n(n-1)/2}.$$

Exercise 6.59. Let $n \in \mathbb{N}$. Let u be a column vector with n entries, and let v be a row vector with n entries. (Thus, uv is an $n \times n$-matrix, whereas vu is a 1×1-matrix.) Let A be an $n \times n$-matrix. Prove that

$$\det (A + uv) = \det A + v (\operatorname{adj} A) u$$

(where we regard the 1×1-matrix $v (\operatorname{adj} A) u$ as an element of $\mathbb{K}$).

The next exercise relies on Definition 6.89.

Exercise 6.60. Let $n \in \mathbb{N}$. Let $u \in \mathbb{K}^{n \times 1}$ be a column vector with n entries, and let $v \in \mathbb{K}^{1 \times n}$ be a row vector with n entries. (Thus, uv is an $n \times n$-matrix, whereas vu is a 1×1-matrix.) Let $h \in \mathbb{K}$. Let H be the 1×1-matrix $\left(\begin{array}{c} h \end{array} \right) \in \mathbb{K}^{1 \times 1}$.

(a) Prove that every $n \times n$-matrix $A \in \mathbb{K}^{n \times n}$ satisfies

$$\det \left(\begin{array}{cc} A & u \\ v & H \end{array} \right) = h \det A - v (\operatorname{adj} A) u$$

(where we regard the 1×1-matrix $v (\operatorname{adj} A) u$ as an element of $\mathbb{K}$).

(b) Write the vector u in the form $u = (u_1, u_2, \ldots, u_n)^T$. Write the vector v in the form $v = (v_1, v_2, \ldots, v_n)$.

For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

Let $d_1, d_2, \ldots, d_n$ be n elements of $\mathbb{K}$. Let D be the $n \times n$-matrix $\left(d_i \delta_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Prove that

$$\det \left(\begin{array}{cc} D & u \\ v & H \end{array} \right) = h \cdot (d_1 d_2 \cdots d_n) - \sum_{i=1}^{n} u_i v_i \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq i}} d_j .$$

Example 6.168. Let $n = 3$. Then, Exercise 6.60 **(b)** states that

$$\det \begin{pmatrix} d_1 & 0 & 0 & u_1 \\ 0 & d_2 & 0 & u_2 \\ 0 & 0 & d_3 & u_3 \\ v_1 & v_2 & v_3 & h \end{pmatrix} = h \cdot (d_1 d_2 d_3) - \sum_{i=1}^{3} u_i v_i \prod_{\substack{j \in \{1,2,3\}; \\ j \neq i}} d_j$$
$$= h d_1 d_2 d_3 - (u_1 v_1 d_2 d_3 + u_2 v_2 d_1 d_3 + u_3 v_3 d_1 d_2)$$

for any ten elements $u_1, u_2, u_3, v_1, v_2, v_3, d_1, d_2, d_3, h$ of $\mathbb{K}$.

Exercise 6.61. Let $P = \sum_{k=0}^{d} p_k X^k$ and $Q = \sum_{k=0}^{e} q_k X^k$ be two polynomials over $\mathbb{K}$ (where $p_0, p_1, \ldots, p_d \in \mathbb{K}$ and $q_0, q_1, \ldots, q_e \in \mathbb{K}$ are their coefficients) such that $d + e > 0$. Define a $(d+e) \times (d+e)$-matrix A as follows:

- For every $k \in \{1, 2, \ldots, e\}$, the k-th row of A is

$$\left(\underbrace{0, 0, \ldots, 0}_{k-1 \text{ zeroes}}, p_d, p_{d-1}, \ldots, p_1, p_0, \underbrace{0, 0, \ldots, 0}_{e-k \text{ zeroes}} \right).$$

- For every $k \in \{1, 2, \ldots, d\}$, the $(e+k)$-th row of A is

$$\left(\underbrace{0, 0, \ldots, 0}_{k-1 \text{ zeroes}}, q_e, q_{e-1}, \ldots, q_1, q_0, \underbrace{0, 0, \ldots, 0}_{d-k \text{ zeroes}} \right).$$

(For example, if $d = 4$ and $e = 3$, then

$$A = \begin{pmatrix} p_4 & p_3 & p_2 & p_1 & p_0 & 0 & 0 \\ 0 & p_4 & p_3 & p_2 & p_1 & p_0 & 0 \\ 0 & 0 & p_4 & p_3 & p_2 & p_1 & p_0 \\ q_3 & q_2 & q_1 & q_0 & 0 & 0 & 0 \\ 0 & q_3 & q_2 & q_1 & q_0 & 0 & 0 \\ 0 & 0 & q_3 & q_2 & q_1 & q_0 & 0 \\ 0 & 0 & 0 & q_3 & q_2 & q_1 & q_0 \end{pmatrix}.$$

)

Assume that the polynomials P and Q have a common root z (that is, there exists a $z \in \mathbb{K}$ such that $P(z) = 0$ and $Q(z) = 0$). Show that $\det A = 0$.

[**Hint:** Find a column vector v with $d + e$ entries satisfying $Av = 0_{(d+e) \times 1}$; then apply Corollary 6.102.]

Remark 6.169. The matrix A in Exercise 6.61 is called the *Sylvester matrix* of the polynomials P and Q (for degrees d and e); its determinant $\det A$ is known as their *resultant* (at least when d and e are actually the degrees of P and Q). According to the exercise, the condition $\det A = 0$ is necessary for P and Q to have a common root. In the general case, the converse does not hold: For one, you can always force $\det A$ to be 0 by taking $d > \deg P$ and $e > \deg Q$ (so $p_d = 0$ and $q_e = 0$, and thus the 1-st column of A consists of zeroes). More importantly, the resultant of the two polynomials $X^3 - 1$ and $X^2 + X + 1$ is 0, but they only have common roots in $\mathbb{C}$, not in $\mathbb{R}$. Thus, there is more to common roots than just the vanishing of a determinant.

However, if $\mathbb{K}$ is an algebraically closed field (I won't go into the details of what this means, but an example of such a field is $\mathbb{C}$), and if $d = \deg P$ and $e = \deg Q$, then the polynomials P and Q have a common root **if and only if** their resultant is 0.

7. Solutions

This section contains solutions (or, sometimes, solution sketches) to some of the exercises in the text, as well as occasional remarks. I do not recommend reading them before trying to solve the problem on your own.

7.1. Solution to Exercise 1.1

Proof of Lemma 1.3. **(a)** Let S be a subset of U. We must prove that $|f(S)| \leq |S|$.

Let $(s_1, s_2, \ldots, s_k)$ be a list of all elements of S (with no repetitions).[283] Thus, $\{s_1, s_2, \ldots, s_k\} = S$ and $k = |S|$. Now,

$$f\left(\underbrace{S}_{=\{s_1,s_2,\ldots,s_k\}}\right) = f(\{s_1, s_2, \ldots, s_k\}) = \{f(s_1), f(s_2), \ldots, f(s_k)\}.$$

Hence,

$$\begin{aligned} |f(S)| &= |\{f(s_1), f(s_2), \ldots, f(s_k)\}| \\ &\leq k \qquad \left(\begin{array}{c}\text{since there are at most } k \text{ distinct elements} \\ \text{among } f(s_1), f(s_2), \ldots, f(s_k)\end{array}\right) \\ &= |S|. \end{aligned}$$

This proves Lemma 1.3 **(a)**.

(b) Let u and v be two elements of U satisfying $f(u) = f(v)$. We shall prove that $u = v$.

[283] Such a list exists, since the set S is finite.

Indeed, assume the contrary. Thus, $u \neq v$. Hence, $u \in U \setminus \{v\}$. But $v \in U$ and therefore $|U \setminus \{v\}| = |U| - 1$.

Lemma 1.3 **(a)** (applied to $S = U \setminus \{v\}$) shows that $|f(U \setminus \{v\})| \leq |U \setminus \{v\}| = |U| - 1$.

Next, I claim that $q \in f(U \setminus \{v\})$ for each $q \in f(U)$.

Indeed, let $q \in f(U)$ be arbitrary. We want to show that $q \in f(U \setminus \{v\})$.

We know that $q \in f(U)$. Hence, there exists some $p \in U$ satisfying $q = f(p)$. Consider this p. If $p = v$, then

$$q = f\left(\underbrace{p}_{=v}\right) = f(v) = f\left(\underbrace{u}_{\in U \setminus \{v\}}\right) \in f(U \setminus \{v\}).$$

Hence, if $p = v$, then $q \in f(U \setminus \{v\})$ is proven. Thus, for the rest of the proof of $q \in f(U \setminus \{v\})$, we WLOG assume that $p \neq v$.

Hence, $p \in U \setminus \{v\}$. Now, $q = f\left(\underbrace{p}_{\in U \setminus \{v\}}\right) \in f(U \setminus \{v\})$. This completes the proof of $q \in f(U \setminus \{v\})$.

Now, forget that we fixed q. We thus have shown that $q \in f(U \setminus \{v\})$ for each $q \in f(U)$. In other words, $f(U) \subseteq f(U \setminus \{v\})$. Hence, $|f(U)| \leq |f(U \setminus \{v\})| \leq |U| - 1 < |U|$. Thus, $|U| > |f(U)| \geq |U|$. This is absurd. This contradiction shows that our assumption was false. Thus, $u = v$ is proven.

Now, forget that we fixed u and v. We thus have shown that if u and v are two elements of U satisfying $f(u) = f(v)$, then $u = v$. In other words, the map f is injective. This proves Lemma 1.3 **(b)**.

(c) Assume that f is injective. Let S be a subset of U. We must prove that $|f(S)| = |S|$.

Let $(s_1, s_2, \ldots, s_k)$ be a list of all elements of S (with no repetitions).[284] Thus, $\{s_1, s_2, \ldots, s_k\} = S$ and $k = |S|$. Furthermore, the elements $s_1, s_2, \ldots, s_k$ are pairwise distinct (since $(s_1, s_2, \ldots, s_k)$ is a list with no repetitions). In other words,

$$s_i \neq s_j \qquad \text{for any two distinct elements } i \text{ and } j \text{ of } \{1, 2, \ldots, k\}.$$

Thus,

$$f(s_i) \neq f(s_j) \qquad \text{for any two distinct elements } i \text{ and } j \text{ of } \{1, 2, \ldots, k\}$$

(since the map f is injective, and thus $f(s_i) \neq f(s_j)$ follows from $s_i \neq s_j$). In other words, the k elements $f(s_1), f(s_2), \ldots, f(s_k)$ are pairwise distinct. Hence, $|\{f(s_1), f(s_2), \ldots, f(s_k)\}| = k$.

Now,

$$f\left(\underbrace{S}_{=\{s_1, s_2, \ldots, s_k\}}\right) = f(\{s_1, s_2, \ldots, s_k\}) = \{f(s_1), f(s_2), \ldots, f(s_k)\}.$$

[284] Such a list exists, since the set S is finite.

Hence,

$$\left|\underbrace{f(S)}_{=\{f(s_1),f(s_2),\ldots,f(s_k)\}}\right| = |\{f(s_1), f(s_2), \ldots, f(s_k)\}| = k = |S|.$$

This proves Lemma 1.3 **(c)**. □

Proof of Lemma 1.4. Assume that f is surjective. Thus, $f(U) = V$. Hence, $|f(U)| = |V| \geq |U|$ (since $|U| \leq |V|$). Hence, Lemma 1.3 **(b)** shows that the map f is injective. Since f is both surjective and injective, we see that f is bijective.

Now, forget that we have assumed that f is surjective. We thus have shown that if f is surjective, then f is bijective. Of course, the converse also holds: If f is bijective, then f is surjective. Hence, f is surjective if and only if f is bijective. This proves Lemma 1.4. □

Proof of Lemma 1.5. Assume that f is injective. Thus, Lemma 1.3 **(c)** (applied to $S = U$) yields $|f(U)| = |U|$ (since U is a subset of U). Thus, $|f(U)| = |U| \geq |V|$.

But the following fact is well-known: If P is a finite set, and if Q is a subset of P such that $|Q| \geq |P|$, then $Q = P$. Applying this to $P = V$ and $Q = f(U)$, we conclude that $f(U) = V$ (since $f(U)$ is a subset of V such that $|f(U)| \geq |V|$). In other words, the map f is surjective. Since f is both surjective and injective, we see that f is bijective.

Now, forget that we have assumed that f is injective. We thus have shown that if f is injective, then f is bijective. Of course, the converse also holds: If f is bijective, then f is injective. Hence, f is injective if and only if f is bijective. This proves Lemma 1.5. □

Solution to Exercise 1.1. We have proven Lemma 1.3, Lemma 1.4 and Lemma 1.5. Thus, Exercise 1.1 is solved. □

7.2. Solution to Exercise 2.1

Solution to Exercise 2.1. First, we notice that the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields

$$b_2 = \frac{b_{2-1}^2 + q}{b_{2-2}} = \frac{b_1^2 + q}{b_0} = \frac{1^2 + q}{1} \qquad \text{(since } b_0 = 1 \text{ and } b_1 = 1\text{)}$$
$$= q + 1.$$

Comparing this with

$$(q+2)\underbrace{b_{2-1}}_{=b_1=1} - \underbrace{b_{2-2}}_{=b_0=1} = (q+2) \cdot 1 - 1 = q + 1,$$

we obtain $b_2 = (q+2) b_{2-1} - b_{2-2}$.

(a) We shall prove Exercise 2.1 **(a)** by induction on n starting at 2:

Induction base: We have already shown that $b_2 = (q+2)\, b_{2-1} - b_{2-2}$. In other words, Exercise 2.1 **(a)** holds for $n = 2$. This completes the induction base.

Induction step: Let $m \in \mathbb{Z}_{\geq 2}$. Assume that Exercise 2.1 **(a)** holds for $n = m$. We must prove that Exercise 2.1 **(a)** holds for $n = m+1$.

We have assumed that Exercise 2.1 **(a)** holds for $n = m$. In other words, we have $b_m = (q+2)\, b_{m-1} - b_{m-2}$. Thus,

$$b_m - (q+2)\, b_{m-1} = -b_{m-2}. \tag{477}$$

We have $m \in \mathbb{Z}_{\geq 2}$. Thus, m is an integer that is ≥ 2. Hence, the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields

$$b_m = \frac{b_{m-1}^2 + q}{b_{m-2}}.$$

Multiplying this equality by b_{m-2}, we obtain

$$b_m b_{m-2} = b_{m-1}^2 + q. \tag{478}$$

Also, $m+1 \geq m \geq 2$ (since m is ≥ 2). Hence, the recursive definition of the sequence $(b_0, b_1, b_2, \ldots)$ yields

$$b_{m+1} = \frac{b_{(m+1)-1}^2 + q}{b_{(m+1)-2}} = \frac{b_m^2 + q}{b_{m-1}}.$$

Multiplying this equality by b_{m-1}, we obtain $b_{m-1} b_{m+1} = b_m^2 + q$. Now,

$$\begin{aligned} b_{m-1}\left(b_{m+1} - (q+2)\, b_m + b_{m-1}\right) &= \underbrace{b_{m-1} b_{m+1}}_{= b_m^2 + q} - (q+2)\, b_{m-1} b_m + b_{m-1}^2 \\ &= b_m^2 + q - (q+2)\, b_{m-1} b_m + b_{m-1}^2 \\ &= b_m \underbrace{\left(b_m - (q+2)\, b_{m-1}\right)}_{\substack{= -b_{m-2} \\ \text{(by (477))}}} + \underbrace{b_{m-1}^2 + q}_{\substack{= b_m b_{m-2} \\ \text{(by (478))}}} \\ &= b_m \left(-b_{m-2}\right) + b_m b_{m-2} = 0. \end{aligned}$$

We can cancel b_{m-1} from this equality (since $b_{m-1} \neq 0$ (because b_{m-1} is a positive rational number)). Thus, we obtain $b_{m+1} - (q+2)\, b_m + b_{m-1} = 0$. Hence,

$$b_{m+1} = (q+2)\, b_m - b_{m-1}.$$

Comparing this with $(q+2) \underbrace{b_{(m+1)-1}}_{= b_m} + \underbrace{b_{(m+1)-2}}_{= b_{m-1}} = (q+2)\, b_m - b_{m-1}$, we obtain $b_{m+1} = (q+2)\, b_{(m+1)-1} + b_{(m+1)-2}$. In other words, Exercise 2.1 **(a)** holds for $n = m+1$. This completes the induction step. Hence, Exercise 2.1 **(a)** is proven by induction.

(b) We shall prove Exercise 2.1 **(b)** by strong induction on n starting at 0:

Induction step: Let $m \in \mathbb{N}$. [285] Assume that Exercise 2.1 **(b)** holds for every $n \in \mathbb{N}$ satisfying $n < m$. We must now show that Exercise 2.1 **(b)** holds for $n = m$.

We have assumed that Exercise 2.1 **(b)** holds for every $n \in \mathbb{N}$ satisfying $n < m$. In other words, we have

$$b_n \in \mathbb{N} \text{ for every } n \in \mathbb{N} \text{ satisfying } n < m. \qquad (479)$$

We must now show that Exercise 2.1 **(b)** holds for $n = m$. In other words, we must show that $b_m \in \mathbb{N}$.

Recall that $(b_0, b_1, b_2, \ldots)$ is a sequence of positive rational numbers. Thus, b_m is a positive rational number.

We are in one of the following three cases:

Case 1: We have $m = 0$.

Case 2: We have $m = 1$.

Case 3: We have $m > 1$.

Let us first consider Case 1. In this case, we have $m = 0$. Thus, $b_m = b_0 = 1 \in \mathbb{N}$. Hence, $b_m \in \mathbb{N}$ is proven in Case 1.

Similarly, we can prove $b_m \in \mathbb{N}$ in Case 2 (using $b_1 = 1$). It thus remains to prove $b_m \in \mathbb{N}$ in Case 3.

So let us consider Case 3. In this case, we have $m > 1$. Thus, $m \geq 2$ (since m is an integer), so that $m \in \mathbb{Z}_{\geq 2}$. Thus, Exercise 2.1 **(a)** (applied to $n = m$) yields $b_m = (q+2) b_{m-1} - b_{m-2}$.

But $m \geq 2 \geq 1$, so that $m - 1 \in \mathbb{N}$ and $m - 1 < m$. Hence, (479) (applied to $n = m - 1$) yields $b_{m-1} \in \mathbb{N} \subseteq \mathbb{Z}$.

Also, $m \geq 2$, so that $m - 2 \in \mathbb{N}$ and $m - 2 < m$. Hence, (479) (applied to $n = m - 2$) yields $b_{m-2} \in \mathbb{N} \subseteq \mathbb{Z}$.

So we know that b_{m-1} and b_{m-2} are both integers (since $b_{m-2} \in \mathbb{Z}$ and $b_{m-1} \in \mathbb{Z}$). Hence, $(q+2) b_{m-1} - b_{m-2}$ is an integer as well (since q is an integer). In other words, b_m is an integer (because $b_m = (q+2) b_{m-1} - b_{m-2}$). Since b_m is positive, we thus conclude that b_m is a positive integer. Hence, $b_m \in \mathbb{N}$. This shows that $b_m \in \mathbb{N}$ in Case 3.

We now have proven $b_m \in \mathbb{N}$ in each of the three Cases 1, 2 and 3. Since these three Cases cover all possibilities, we thus conclude that $b_m \in \mathbb{N}$ always holds. In other words, Exercise 2.1 **(b)** holds for $n = m$. This completes the induction step. Thus, Exercise 2.1 **(b)** is proven by strong induction. □

7.3. Solution to Exercise 2.2

Proof of Proposition 2.102. We shall prove Proposition 2.102 by induction on n:

Induction base: We want to prove that Proposition 2.102 holds when $n = 0$. In other words, we want to prove the following claim:

[285] In order to match the notations used in Theorem 2.60, we should be saying "Let $m \in \mathbb{Z}_{\geq 0}$" here, rather than "Let $m \in \mathbb{N}$". But of course, this amounts to the same thing, since $\mathbb{N} = \mathbb{Z}_{\geq 0}$.

> *Claim 1:* Let $X_1, X_2, \ldots, X_{0+1}$ be $0+1$ sets. For each $i \in \{1, 2, \ldots, 0\}$, let $f_i : X_i \to X_{i+1}$ be an invertible map. Then, the map $f_0 \circ f_{0-1} \circ \cdots \circ f_1 : X_1 \to X_{0+1}$ is invertible as well, and its inverse is
>
> $$(f_0 \circ f_{0-1} \circ \cdots \circ f_1)^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_0^{-1}.$$

But this is straightforward:

[*Proof of Claim 1:* The equality (138) (applied to $n = 0$) yields $f_0 \circ f_{0-1} \circ \cdots \circ f_1 = \mathrm{id}_{X_1}$. Similarly,

$$f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_0^{-1} = \mathrm{id}_{X_1}. \tag{480}$$

Now, $f_0 \circ f_{0-1} \circ \cdots \circ f_1 = \mathrm{id}_{X_1}$ and $X_{0+1} = X_1$. Hence, the map $f_0 \circ f_{0-1} \circ \cdots \circ f_1 : X_1 \to X_{0+1}$ is the same as the map $\mathrm{id}_{X_1} : X_1 \to X_1$, and therefore is invertible (because the map $\mathrm{id}_{X_1} : X_1 \to X_1$ clearly is invertible). Moreover, its inverse is

$$\Bigg(\underbrace{f_0 \circ f_{0-1} \circ \cdots \circ f_1}_{=\mathrm{id}_{X_1}}\Bigg)^{-1} = (\mathrm{id}_{X_1})^{-1} = \mathrm{id}_{X_1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_0^{-1} \qquad \text{(by (480))}.$$

This completes the proof of Claim 1.]

We have now proven Claim 1. In other words, Proposition 2.102 holds when $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Proposition 2.102 holds when $n = m$. We must now prove that Proposition 2.102 holds when $n = m + 1$.

We have assumed that Proposition 2.102 holds when $n = m$. In other words, the following claim holds:

> *Claim 2:* Let $X_1, X_2, \ldots, X_{m+1}$ be $m+1$ sets. For each $i \in \{1, 2, \ldots, m\}$, let $f_i : X_i \to X_{i+1}$ be an invertible map. Then, the map $f_m \circ f_{m-1} \circ \cdots \circ f_1 : X_1 \to X_{m+1}$ is invertible as well, and its inverse is
>
> $$(f_m \circ f_{m-1} \circ \cdots \circ f_1)^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_m^{-1}.$$

We must now prove that Proposition 2.102 holds when $n = m + 1$. In other words, we must prove the following claim:

> *Claim 3:* Let $X_1, X_2, \ldots, X_{(m+1)+1}$ be $(m+1)+1$ sets. For each $i \in \{1, 2, \ldots, m+1\}$, let $f_i : X_i \to X_{i+1}$ be an invertible map. Then, the map $f_{m+1} \circ f_{(m+1)-1} \circ \cdots \circ f_1 : X_1 \to X_{(m+1)+1}$ is invertible as well, and its inverse is
>
> $$\left(f_{m+1} \circ f_{(m+1)-1} \circ \cdots \circ f_1\right)^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_{m+1}^{-1}.$$

[*Proof of Claim 3:* We have $m \in \mathbb{N}$; thus, $m+1$ is a positive integer. Hence, $m+1 \in \{1,2,\ldots,m+1\}$ and $m+1 \geq 1$. Thus, Theorem 2.95 **(b)** (applied to $n = m+1$) yields

$$f_{m+1} \circ f_{(m+1)-1} \circ \cdots \circ f_1 = f_{m+1} \circ \underbrace{\left(f_{(m+1)-1} \circ f_{(m+1)-2} \circ \cdots \circ f_1\right)}_{=f_m \circ f_{m-1} \circ \cdots \circ f_1} = f_{m+1} \circ (f_m \circ f_{m-1} \circ \cdots \circ f_1). \qquad (481)$$

Also, Theorem 2.95 **(c)** (applied to $m+1$, X_{m+3-i} and f_{m+2-i}^{-1} instead of n, X_i and f_i) yields

$$f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_{m+1}^{-1} = \left(f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_m^{-1}\right) \circ f_{m+1}^{-1}. \qquad (482)$$

Claim 2 yields that the map $f_m \circ f_{m-1} \circ \cdots \circ f_1 : X_1 \to X_{m+1}$ is invertible, and that its inverse is

$$(f_m \circ f_{m-1} \circ \cdots \circ f_1)^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_m^{-1}.$$

Hence,

$$\begin{aligned} &\underbrace{(f_m \circ f_{m-1} \circ \cdots \circ f_1)^{-1}}_{=f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_m^{-1}} \circ f_{m+1}^{-1} \\ &= \left(f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_m^{-1}\right) \circ f_{m+1}^{-1} \\ &= f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_{m+1}^{-1} \qquad \text{(by (482))}. \end{aligned} \qquad (483)$$

We know (by our assumption) that for each $i \in \{1,2,\ldots,m+1\}$, the map $f_i : X_i \to X_{i+1}$ is an invertible map. Applying this to $i = m+1$, we conclude that $f_{m+1} : X_{m+1} \to X_{(m+1)+1}$ is an invertible map (since $m+1 \in \{1,2,\ldots,m+1\}$).

Now, we know that $f_m \circ f_{m-1} \circ \cdots \circ f_1 : X_1 \to X_{m+1}$ and $f_{m+1} : X_{m+1} \to X_{(m+1)+1}$ are two invertible maps. Hence, Proposition 2.101 (applied to $X = X_1$, $Y = X_{m+1}$, $Z = X_{(m+1)+1}$, $b = f_m \circ f_{m-1} \circ \cdots \circ f_1$ and $a = f_{m+1}$) yields that the map $f_{m+1} \circ (f_m \circ f_{m-1} \circ \cdots \circ f_1) : X_1 \to X_{(m+1)+1}$ is invertible as well, and that its inverse is

$$(f_{m+1} \circ (f_m \circ f_{m-1} \circ \cdots \circ f_1))^{-1} = (f_m \circ f_{m-1} \circ \cdots \circ f_1)^{-1} \circ f_{m+1}^{-1}.$$

In view of

$$f_{m+1} \circ (f_m \circ f_{m-1} \circ \cdots \circ f_1) = f_{m+1} \circ f_{(m+1)-1} \circ \cdots \circ f_1 \qquad \text{(by (481))}$$

and

$$(f_m \circ f_{m-1} \circ \cdots \circ f_1)^{-1} \circ f_{m+1}^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_{m+1}^{-1} \qquad \text{(by (483))},$$

this rewrites as follows: The map $f_{m+1} \circ f_{(m+1)-1} \circ \cdots \circ f_1 : X_1 \to X_{(m+1)+1}$ is invertible as well, and its inverse is

$$\left(f_{m+1} \circ f_{(m+1)-1} \circ \cdots \circ f_1\right)^{-1} = f_1^{-1} \circ f_2^{-1} \circ \cdots \circ f_{m+1}^{-1}.$$

This proves Claim 3.]

We have now proven Claim 3. In other words, Proposition 2.102 holds when $n = m + 1$. This completes the induction step. Hence, Proposition 2.102 is proven by induction. □

Solution to Exercise 2.2. We have proven Proposition 2.102. Thus, Exercise 2.2 is solved. □

7.4. Solution to Exercise 2.3

Let us first prove a lemma:

Lemma 7.1. Let S be any set. Let $(a_s)_{s\in S}$ be an $\mathbb{A}$-valued S-family. Let T_1 and T_2 be two finite subsets T of S such that (206) holds. Then,

$$\sum_{s\in T_1} a_s = \sum_{s\in T_2} a_s.$$

(Note that we are not using the notation introduced in Definition 2.144 yet, because we have not proven that this notation is well-defined.)

Proof of Lemma 7.1. We know that T_1 is a finite subset T of S such that (206) holds. Hence, T_1 is a finite subset of S.

We know that T_2 is a finite subset T of S such that (206) holds. In other words, T_2 is a finite subset of S and has the property that (206) holds for $T = T_2$.

In particular, (206) holds for $T = T_2$. In other words,

$$\text{every } s \in S \setminus T_2 \text{ satisfies } a_s = 0. \qquad (484)$$

The set $T_1 \setminus T_2$ is a subset of the finite set T_1, and thus is finite.

If $s \in T_1 \setminus T_2$, then $s \in \underbrace{T_1}_{\subseteq S} \setminus T_2 \subseteq S \setminus T_2$ and thus

$$a_s = 0 \qquad (485)$$

(by (484)). Now,

$$\sum_{s\in T_1\setminus T_2} \underbrace{a_s}_{\substack{=0 \\ \text{(by (485))}}} = \sum_{s\in T_1\setminus T_2} 0 = 0$$

(by Theorem 2.126 (applied to $T_1 \setminus T_2$ instead of S)).

It is a straightforward exercise in set theory to show that if A and B are any two sets, then $A \cap B$ and $A \setminus B$ are two subsets of A satisfying

$$(A \cap B) \cap (A \setminus B) = \varnothing \qquad \text{and} \qquad (A \cap B) \cup (A \setminus B) = A.$$

Applying this to $A = T_1$ and $B = T_2$, we conclude that $T_1 \cap T_2$ and $T_1 \setminus T_2$ are two subsets of T_1 satisfying

$$(T_1 \cap T_2) \cap (T_1 \setminus T_2) = \varnothing \qquad \text{and} \qquad (T_1 \cap T_2) \cup (T_1 \setminus T_2) = T_1.$$

Hence, Theorem 2.130 (applied to T_1, $T_1 \cap T_2$ and $T_1 \setminus T_2$ instead of S, X and Y) yields

$$\sum_{s \in T_1} a_s = \sum_{s \in T_1 \cap T_2} a_s + \underbrace{\sum_{s \in T_1 \setminus T_2} a_s}_{=0} = \sum_{s \in T_1 \cap T_2} a_s. \tag{486}$$

The same argument (with the roles of T_1 and T_2 interchanged) yields

$$\sum_{s \in T_2} a_s = \sum_{s \in T_2 \cap T_1} a_s = \sum_{s \in T_1 \cap T_2} a_s$$

(since $T_2 \cap T_1 = T_1 \cap T_2$). Comparing this equality with (486), we obtain $\sum\limits_{s \in T_1} a_s = \sum\limits_{s \in T_2} a_s$. This proves Lemma 7.1. □

Proof of Proposition 2.145. We shall **not** use the notation introduced in Definition 2.144 in this proof, because we have not yet convinced ourselves that this notation is well-defined.

(a) Lemma 7.1 shows that if T_1 and T_2 are two finite subsets T of S such that (206) holds, then

$$\sum_{s \in T_1} a_s = \sum_{s \in T_2} a_s.$$

In other words, if T is a finite subset of S such that (206) holds, then the sum $\sum\limits_{s \in T} a_s$ does not depend on the choice of T. This proves Proposition 2.145 **(a)**.

(b) Assume that the set S is finite. Note that every $s \in S \setminus S$ satisfies $a_s = 0$. (Indeed, this is vacuously true, because there exists no $s \in S \setminus S$ (since $S \setminus S = \varnothing$).)

Let T be a finite subset of S such that

$$\text{every } s \in S \setminus T \text{ satisfies } a_s = 0. \tag{487}$$

(It is easy to see that such a subset T exists[286].)

We shall prove the equality $\sum\limits_{s \in S} a_s = \sum\limits_{s \in T} a_s$. (Note that both sums in this equality are defined according to Definition 2.111, not according to Definition 2.144.)

Indeed, we have

$$\sum_{s \in S \setminus T} \underbrace{a_s}_{\substack{=0 \\ \text{(by (487))}}} = \sum_{s \in S \setminus T} 0 = 0$$

(by Theorem 2.126 (applied to $S \setminus T$ instead of S)).

[286] *Proof.* Indeed, S itself is such a subset T (because S is a finite subset of S and has the property that every $s \in S \setminus S$ satisfies $a_s = 0$).

Now, T and $S \setminus T$ are two subsets of S. A well-known fact from set theory says that if A is a subset of a set B, then

$$A \cap (B \setminus A) = \varnothing \qquad \text{and} \qquad A \cup (B \setminus A) = B.$$

We can apply this to $A = T$ and $B = S$. We thus obtain

$$T \cap (S \setminus T) = \varnothing \qquad \text{and} \qquad T \cup (S \setminus T) = S.$$

Thus, Theorem 2.130 (applied to T and $S \setminus T$ instead of X and Y) yields

$$\sum_{s \in S} a_s = \sum_{s \in T} a_s + \underbrace{\sum_{s \in S \setminus T} a_s}_{=0} = \sum_{s \in T} a_s. \tag{488}$$

Now, recall that T is a finite subset of S such that every $s \in S \setminus T$ satisfies $a_s = 0$. Hence, Definition 2.144 defines the sum $\sum_{s \in S} a_s$ to be $\sum_{s \in T} a_s$. Thus,

$$\begin{aligned} &\left(\text{the sum } \sum_{s \in S} a_s \text{ defined in Definition 2.144}\right) \\ &= \sum_{s \in T} a_s = \sum_{s \in S} a_s \qquad \text{(by (488))} \\ &= \left(\text{the sum } \sum_{s \in S} a_s \text{ defined in Definition 2.111}\right). \end{aligned}$$

In other words, the sum $\sum_{s \in S} a_s$ defined in Definition 2.144 is identical with the sum $\sum_{s \in S} a_s$ defined in Definition 2.111. This proves Proposition 2.145 **(b)**. □

Now that Proposition 2.145 has been proven, we can use Definition 2.144, and we shall do so in the remainder of this section.

Let us state a takeaway from Definition 2.144:

Corollary 7.2. Let S be any set. Let $(a_s)_{s \in S}$ be a finitely supported $\mathbb{A}$-valued S-family. Let T be a finite subset of S such that

$$\text{every } s \in S \setminus T \text{ satisfies } a_s = 0.$$

Then,

$$\sum_{s \in S} a_s = \sum_{s \in T} a_s.$$

Proof of Corollary 7.2. Recall that T is a finite subset of S such that every $s \in S \setminus T$ satisfies $a_s = 0$. Hence, Definition 2.144 defines the sum $\sum_{s \in S} a_s$ to be $\sum_{s \in T} a_s$. Thus, $\sum_{s \in S} a_s = \sum_{s \in T} a_s$. This proves Corollary 7.2. □

Next, let us generalize Corollary 7.2, no longer requiring T to be finite:

Corollary 7.3. Let S be any set. Let $(a_s)_{s\in S}$ be a finitely supported $\mathbb{A}$-valued S-family. Let T be a subset of S such that

$$\text{every } s \in S \setminus T \text{ satisfies } a_s = 0. \tag{489}$$

Then, the $\mathbb{A}$-valued T-family $(a_s)_{s\in T}$ is finitely supported as well, and satisfies

$$\sum_{s\in S} a_s = \sum_{s\in T} a_s. \tag{490}$$

Note that both sums $\sum\limits_{s\in S} a_s$ and $\sum\limits_{s\in T} a_s$ in Corollary 7.3 are defined according to Definition 2.144.

Proof of Corollary 7.3. We have assumed that the S-family $(a_s)_{s\in S}$ is finitely supported. In other words, only finitely many $s \in S$ satisfy $a_s \neq 0$ (by the definition of "finitely supported"). In other words, there exists a finite subset Q of S such that

$$\text{every } s \in S \setminus Q \text{ satisfies } a_s = 0. \tag{491}$$

Consider this Q.

The set $T \cap Q$ is a subset of Q, and thus is finite (since Q is finite). Also, $T \cap Q$ is a subset of T. Moreover, every $s \in T \setminus (T \cap Q)$ satisfies $a_s = 0$ [287]. Hence, there exists a finite subset F of T such that every $s \in T \setminus F$ satisfies $a_s = 0$ (namely, $F = T \cap Q$). In other words, only finitely many $s \in T$ satisfy $a_s \neq 0$. In other words, the $\mathbb{A}$-valued T-family $(a_s)_{s\in T}$ is finitely supported (by the definition of "finitely supported"). It now remains to prove (490).

The set $T \cap Q$ is finite and is a subset of S (since $T \cap Q \subseteq Q \subseteq S$), and has the property that

$$\text{every } s \in S \setminus (T \cap Q) \text{ satisfies } a_s = 0. \tag{492}$$

[*Proof of (492):* Let $s \in S \setminus (T \cap Q)$. We must prove that $a_s = 0$.

We have $s \in S \setminus (T \cap Q)$. In other words, $s \in S$ and $s \notin T \cap Q$.

We are in one of the following two cases:

Case 1: We have $s \in Q$.

Case 2: We don't have $s \in Q$.

Let us first consider Case 1. In this case, we have $s \in Q$. Combining $s \in Q$ with $s \notin T \cap Q$, we obtain $s \in Q \setminus (T \cap Q) = \underbrace{Q}_{\subseteq S} \setminus T \subseteq S \setminus T$. Thus, $a_s = 0$ (by (489)). Hence, $a_s = 0$ is proven in Case 1.

[287] *Proof.* Let $s \in T \setminus (T \cap Q)$. We must prove that $a_s = 0$.

We have $s \in T \setminus (T \cap Q) = \underbrace{T}_{\subseteq S} \setminus Q \subseteq S \setminus Q$. Hence, (491) shows that $a_s = 0$. Qed.

Let us now consider Case 2. In this case, we don't have $s \in Q$. Hence, $s \notin Q$. Combining $s \in S$ with $s \notin Q$, we obtain $s \in S \setminus Q$. Hence, (491) yields $a_s = 0$. Thus, $a_s = 0$ is proven in Case 2.

We now have shown that $a_s = 0$ in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that $a_s = 0$ always holds. This proves (492).]

Hence, Corollary 7.2 (applied to $T \cap Q$ instead of T) yields

$$\sum_{s \in S} a_s = \sum_{s \in T \cap Q} a_s. \tag{493}$$

On the other hand, $T \cap Q$ is a finite subset of T, and has the property that

$$\text{every } s \in T \setminus (T \cap Q) \text{ satisfies } a_s = 0. \tag{494}$$

(Indeed, this follows from (491), because every $s \in T \setminus (T \cap Q)$ satisfies $s \in T \setminus (T \cap Q) = \underbrace{T}_{\subseteq S} \setminus Q \subseteq S \setminus Q$.)

Hence, Corollary 7.2 (applied to $T \cap Q$ and T instead of T and S) yields

$$\sum_{s \in T} a_s = \sum_{s \in T \cap Q} a_s$$

(since the $\mathbb{A}$-valued T-family $(a_s)_{s \in T}$ is finitely supported). Comparing this with (493), we obtain $\sum_{s \in S} a_s = \sum_{s \in T} a_s$. This proves (490). Hence, Corollary 7.3 is proven. □

Proof of Theorem 2.146. The S-family $(a_s)_{s \in S}$ is finitely supported. In other words, only finitely many $s \in S$ satisfy $a_s \neq 0$ (by the definition of "finitely supported"). In other words, there exists a finite subset A of S such that

$$\text{every } s \in S \setminus A \text{ satisfies } a_s = 0. \tag{495}$$

Consider this A.

The S-family $(b_s)_{s \in S}$ is finitely supported. In other words, only finitely many $s \in S$ satisfy $b_s \neq 0$ (by the definition of "finitely supported"). In other words, there exists a finite subset B of S such that

$$\text{every } s \in S \setminus B \text{ satisfies } b_s = 0. \tag{496}$$

Consider this B.

The set $A \cup B$ is a subset of S (since A and B are subsets of S). Moreover, this set $A \cup B$ is finite (since both sets A and B are finite). Furthermore,

$$\text{every } s \in S \setminus (A \cup B) \text{ satisfies } a_s = 0 \tag{497}$$

[288]. Also,

$$\text{every } s \in S \setminus (A \cup B) \text{ satisfies } b_s = 0 \tag{498}$$

[289]. Thus,

$$\text{every } s \in S \setminus (A \cup B) \text{ satisfies } a_s + b_s = 0 \tag{499}$$

[290]. Thus, there exists a finite subset T of S such that every $s \in S \setminus T$ satisfies $a_s + b_s = 0$ (namely, $T = A \cup B$). In other words, only finitely many $s \in S$ satisfy $a_s + b_s \neq 0$. In other words, the S-family $(a_s + b_s)_{s \in S}$ is finitely supported.

The set $A \cup B$ is finite. Hence, Theorem 2.122 (applied to $A \cup B$ instead of S) yields

$$\sum_{s \in A \cup B} (a_s + b_s) = \sum_{s \in A \cup B} a_s + \sum_{s \in A \cup B} b_s. \tag{500}$$

We know that $A \cup B$ is a finite subset of S such that every $s \in S \setminus (A \cup B)$ satisfies $a_s = 0$ (by (497)). Thus, Corollary 7.2 (applied to $A \cup B$ instead of T) yields

$$\sum_{s \in S} a_s = \sum_{s \in A \cup B} a_s. \tag{501}$$

We know that $A \cup B$ is a finite subset of S such that every $s \in S \setminus (A \cup B)$ satisfies $b_s = 0$ (by (498)). Thus, Corollary 7.2 (applied to $(b_s)_{s \in S}$ and $A \cup B$ instead of $(a_s)_{s \in S}$ and T) yields

$$\sum_{s \in S} b_s = \sum_{s \in A \cup B} b_s. \tag{502}$$

We know that $A \cup B$ is a finite subset of S such that every $s \in S \setminus (A \cup B)$ satisfies $a_s + b_s = 0$ (by (499)). Thus, Corollary 7.2 (applied to $(a_s + b_s)_{s \in S}$ and $A \cup B$ instead of $(a_s)_{s \in S}$ and T) yields

$$\begin{aligned} \sum_{s \in S} (a_s + b_s) &= \sum_{s \in A \cup B} (a_s + b_s) = \underbrace{\sum_{s \in A \cup B} a_s}_{\substack{= \sum_{s \in S} a_s \\ \text{(by (501))}}} + \underbrace{\sum_{s \in A \cup B} b_s}_{\substack{= \sum_{s \in S} b_s \\ \text{(by (502))}}} \qquad \text{(by (500))} \\ &= \sum_{s \in S} a_s + \sum_{s \in S} b_s. \end{aligned}$$

This completes the proof of Theorem 2.146. □

Before we prove Theorem 2.147, let us restate Corollary 7.2 with different names:

---

[288] *Proof.* Let $s \in S \setminus (A \cup B)$. We must prove that $a_s = 0$.
We have $s \in S \setminus (A \cup B) = (S \setminus A) \setminus B \subseteq S \setminus A$. Thus, (495) yields $a_s = 0$. Qed.

[289] *Proof.* Let $s \in S \setminus (A \cup B)$. We must prove that $b_s = 0$.
We have $s \in S \setminus (A \cup B) = (S \setminus B) \setminus A \subseteq S \setminus B$. Thus, (496) yields $b_s = 0$. Qed.

[290] *Proof.* Every $s \in S \setminus (A \cup B)$ satisfies $\underbrace{a_s}_{\substack{=0 \\ \text{(by (497))}}} + \underbrace{b_s}_{\substack{=0 \\ \text{(by (498))}}} = 0 + 0 = 0$, qed.

Corollary 7.4. Let W be any set. Let $(a_w)_{w\in W}$ be a finitely supported $\mathbb{A}$-valued W-family. Let V be a finite subset of W such that

$$\text{every } w \in W \setminus V \text{ satisfies } a_w = 0.$$

Then,

$$\sum_{w\in W} a_w = \sum_{w\in V} a_w.$$

Proof of Corollary 7.4. Corollary 7.4 is just Corollary 7.2, with all appearances of the letters "S", "s" and "T" replaced by "W", "w" and "V". Thus, Corollary 7.4 follows from the latter corollary. □

Proof of Theorem 2.147. Let V be the subset $f(S)$ of W. (Thus, $V = f(S) = \{f(s) \mid s \in S\}$.) The set $f(S)$ is finite (since the set S is finite). In other words, the set V is finite (since $V = f(S)$).

Define a map $g : S \to V$ by

$$(g(s) = f(s) \text{ for each } s \in S).$$

(This map g is well-defined, since each $s \in S$ satisfies $f(s) \in f(S) = V$.)

For each $w \in W$, we define an element $b_w \in \mathbb{A}$ by

$$b_w = \sum_{\substack{s\in S;\\ f(s)=w}} a_s. \tag{503}$$

Now, each $w \in V$ satisfies

$$\sum_{\substack{s\in S;\\ g(s)=w}} a_s = b_w. \tag{504}$$

[*Proof of (504):* Let $w \in V$. Then, $g(s) = f(s)$ for each $s \in S$ (by the definition of g). Hence, the summation sign "$\sum_{\substack{s\in S;\\ g(s)=w}}$" can be rewritten as "$\sum_{\substack{s\in S;\\ f(s)=w}}$". Thus, $\sum_{\substack{s\in S;\\ g(s)=w}} a_s = \sum_{\substack{s\in S;\\ f(s)=w}} a_s = b_w$ (by (503)). This proves (504).]

Theorem 2.127 (applied to V and g instead of W and f) yields

$$\sum_{s\in S} a_s = \sum_{w\in V} \underbrace{\sum_{\substack{s\in S;\\ g(s)=w}} a_s}_{\substack{=b_w\\ \text{(by (504))}}} = \sum_{w\in V} b_w. \tag{505}$$

Moreover,

$$\text{every } w \in W \setminus V \text{ satisfies } b_w = 0. \tag{506}$$

[*Proof of (506):* Let $w \in W \setminus V$. Thus, $w \in W$ and $w \notin V$. Hence, every $s \in S$ satisfies $f(s) \neq w$ [291]. In other words, there exists no $s \in S$ satisfying $f(s) = w$. Hence, the sum $\sum_{\substack{s \in S;\\ f(s)=w}} a_s$ is an empty sum. Thus, $\sum_{\substack{s \in S;\\ f(s)=w}} a_s = (\text{empty sum}) = 0$. In view of (503), this rewrites as $b_w = 0$. This proves (506).]

Thus, there exists a finite subset T of W such that every $w \in W \setminus T$ satisfies $b_w = 0$ (namely, $T = V$). In other words, only finitely many $w \in W$ satisfy $b_w \neq 0$. In other words, the $\mathbb{A}$-valued W-family $(b_w)_{w \in W}$ is finitely supported. In view of (503), this rewrites as follows: The $\mathbb{A}$-valued W-family $\left(\sum_{\substack{s \in S;\\ f(s)=w}} a_s \right)_{w \in W}$ is finitely supported. Hence, the sum $\sum_{w \in W} \sum_{\substack{s \in S;\\ f(s)=w}} a_s$ is well-defined.

Moreover, Corollary 7.4 (applied to $(b_w)_{w \in W}$ instead of $(a_w)_{w \in W}$) yields that $\sum_{w \in W} b_w = \sum_{w \in V} b_w$. Comparing this with (505), we obtain

$$\sum_{s \in S} a_s = \sum_{w \in W} \underbrace{b_w}_{\substack{= \sum_{\substack{s \in S;\\ f(s)=w}} a_s \\ \text{(by (503))}}} = \sum_{w \in W} \sum_{\substack{s \in S;\\ f(s)=w}} a_s .$$

This completes the proof of Theorem 2.147. □

Proof of Theorem 2.148. The S-family $(a_s)_{s \in S}$ is finitely supported. In other words, only finitely many $s \in S$ satisfy $a_s \neq 0$. In other words, there exists a finite subset T of S such that every $s \in S \setminus T$ satisfies $a_s = 0$. Consider this T, and denote it by P. Thus, P is a finite subset of S and has the property that

$$\text{every } s \in S \setminus P \text{ satisfies } a_s = 0. \tag{507}$$

Thus, Corollary 7.2 (applied to $T = P$) yields

$$\sum_{s \in S} a_s = \sum_{s \in P} a_s . \tag{508}$$

Let $g : P \to W$ be the restriction $f \mid_P$ of the map f to the subset P of S. Hence,

$$g(t) = f(t) \qquad \text{for each } t \in P \tag{509}$$

(by the definition of a restriction of a map).

---

[291] *Proof.* Let $s \in S$. If we had $f(s) = w$, then we would have $w = f\left(\underbrace{s}_{\in S} \right) \in f(S) = V$, which would contradict $w \notin V$. Hence, we cannot have $f(s) = w$. In other words, we have $f(s) \neq w$, qed.

Theorem 2.147 (applied to P and g instead of S and f) yields that the $\mathbb{A}$-valued W-family $\left(\sum\limits_{\substack{s\in P;\\ g(s)=w}} a_s \right)_{w\in W}$ is finitely supported and satisfies

$$\sum_{s\in P} a_s = \sum_{w\in W} \sum_{\substack{s\in P;\\ g(s)=w}} a_s. \tag{510}$$

Hence, (508) becomes

$$\sum_{s\in S} a_s = \sum_{s\in P} a_s = \sum_{w\in W} \sum_{\substack{s\in P;\\ g(s)=w}} a_s. \tag{511}$$

Now, we claim the following:

Claim 1: Let $w \in W$.

(a) The $\mathbb{A}$-valued $\{t \in S \mid f(t) = w\}$-family $(a_s)_{s\in\{t\in S \mid f(t)=w\}}$ is finitely supported (so that the sum $\sum\limits_{\substack{s\in S;\\ f(s)=w}} a_s$ is well-defined).

(b) We have $\sum\limits_{\substack{s\in S;\\ f(s)=w}} a_s = \sum\limits_{\substack{s\in P;\\ g(s)=w}} a_s$.

[*Proof of Claim 1:* Define a subset Q of S by $Q = \{t \in S \mid f(t) = w\}$.

The set $P \cap Q$ is a subset of P, and thus is finite (since P is finite). Moreover, this set $P \cap Q$ is a subset of Q. Finally,

$$\text{every } s \in Q \setminus (P \cap Q) \text{ satisfies } a_s = 0$$

[292]. Hence, there exists a finite subset T of Q such that every $s \in Q \setminus T$ satisfies $a_s = 0$ (namely, $T = P \cap Q$). In other words, only finitely many $s \in Q$ satisfy $a_s \neq 0$. In other words, the $\mathbb{A}$-valued Q-family $(a_s)_{s\in Q}$ is finitely supported. In view of $Q = \{t \in S \mid f(t) = w\}$, this rewrites as follows: The $\mathbb{A}$-valued $\{t \in S \mid f(t) = w\}$-family $(a_s)_{s\in\{t\in S \mid f(t)=w\}}$ is finitely supported. This proves Claim 1 **(a)**.

(b) We know that $P \cap Q$ is a subset of Q such that every $s \in Q \setminus (P \cap Q)$ satisfies $a_s = 0$. Hence, Corollary 7.3 (applied to Q and $P \cap Q$ instead of S and T) shows that the $\mathbb{A}$-valued $P \cap Q$-family $(a_s)_{s\in P\cap Q}$ is finitely supported as well, and satisfies

$$\sum_{s\in Q} a_s = \sum_{s\in P\cap Q} a_s. \tag{512}$$

[292] *Proof:* Let $s \in Q \setminus (P \cap Q)$. We must prove that $a_s = 0$.

We have $s \in Q \setminus (P \cap Q) = \underbrace{Q}_{\subseteq S} \setminus P \subseteq S \setminus P$. Hence, (507) shows that $a_s = 0$. Qed.

But $Q = \{t \in S \mid f(t) = w\}$. Hence, an element t of S belongs to Q if and only if it satisfies $f(t) = w$. In other words, for any element $t \in S$, we have the logical equivalence

$$(t \in Q) \iff (f(t) = w). \tag{513}$$

Now, the definition of the intersection of two sets shows that

$$P \cap Q = \left\{ t \in P \mid \underbrace{t \in Q}_{\substack{\iff (f(t)=w) \\ \text{(by (513)} \\ \text{(since } t \in P \subseteq S))}} \right\} = \left\{ t \in P \mid \underbrace{f(t)}_{\substack{=g(t) \\ \text{(by (509))}}} = w \right\}$$
$$= \{t \in P \mid g(t) = w\}.$$

Now,

$$\underbrace{\sum_{\substack{s \in S; \\ f(s)=w}}}_{\substack{= \sum_{s \in \{t \in S \mid f(t)=w\}} = \sum_{s \in Q} \\ \text{(since } \{t \in S \mid f(t)=w\} = Q)}} a_s = \sum_{s \in Q} a_s = \underbrace{\sum_{s \in P \cap Q}}_{\substack{= \sum_{s \in \{t \in P \mid g(t)=w\}} \\ \text{(since } P \cap Q = \{t \in P \mid g(t)=w\})}} a_s \qquad \text{(by (512))}$$
$$= \underbrace{\sum_{s \in \{t \in P \mid g(t)=w\}}}_{= \sum_{\substack{s \in P; \\ g(s)=w}}} a_s = \sum_{\substack{s \in P; \\ g(s)=w}} a_s.$$

This proves Claim 1 **(b)**.]

Now, Claim 1 **(a)** shows that for each $w \in W$, the $\mathbb{A}$-valued $\{t \in S \mid f(t) = w\}$-family $(a_s)_{s \in \{t \in S \mid f(t)=w\}}$ is finitely supported (so that the sum $\sum_{\substack{s \in S; \\ f(s)=w}} a_s$ is well-defined). Also, Claim 1 **(b)** shows that

$$\sum_{\substack{s \in S; \\ f(s)=w}} a_s = \sum_{\substack{s \in P; \\ g(s)=w}} a_s \qquad \text{for each } w \in W.$$

In other words,

$$\left(\sum_{\substack{s \in S; \\ f(s)=w}} a_s \right)_{w \in W} = \left(\sum_{\substack{s \in P; \\ g(s)=w}} a_s \right)_{w \in W}.$$

Hence, the $\mathbb{A}$-valued W-family $\left(\sum_{\substack{s \in S; \\ f(s)=w}} a_s \right)_{w \in W}$ is finitely supported (since we know that the $\mathbb{A}$-valued W-family $\left(\sum_{\substack{s \in P; \\ g(s)=w}} a_s \right)_{w \in W}$ is finitely supported). Finally, (511) becomes

$$\sum_{s \in S} a_s = \sum_{w \in W} \underbrace{\sum_{\substack{s \in P; \\ g(s)=w}} a_s}_{\substack{= \sum_{\substack{s \in S; \\ f(s)=w}} a_s \\ \text{(by Claim 1 (b))}}} = \sum_{w \in W} \sum_{\substack{s \in S; \\ f(s)=w}} a_s .$$

This completes the proof of Theorem 2.148. □

Solution to Exercise 2.3. We have proven Proposition 2.145, Theorem 2.146, Theorem 2.147 and Theorem 2.148. Thus, Exercise 2.3 is solved. □

7.5. Solution to Exercise 2.4

Proof of Proposition 2.160. **(a)** We have the following chain of logical equivalences:

$$\begin{aligned} & (\text{the integer } n \text{ is even}) \\ \Longleftrightarrow\ & (n \text{ is divisible by } 2) \qquad (\text{by the definition of ``even''}) \\ \Longleftrightarrow\ & (2 \mid n) \iff (2 \mid n - 0) \qquad (\text{since } n = n - 0) \\ \Longleftrightarrow\ & (n \equiv 0 \bmod 2) \qquad (\text{by the definition of ``} n \equiv 0 \bmod 2\text{''}) . \end{aligned}$$

In other words, the integer n is even if and only if $n \equiv 0 \bmod 2$. This proves Proposition 2.160 **(a)**.

(b) Proposition 2.157 **(a)** shows that the integer n is odd if and only if n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$.

Now, we claim the following logical implication:

$$(\text{the integer } n \text{ is odd}) \implies (n \equiv 1 \bmod 2) . \tag{514}$$

[*Proof of (514):* Assume that the integer n is odd. We shall show that $n \equiv 1 \bmod 2$.

Recall that the integer n is odd if and only if n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$. Hence, n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$ (since the integer n is odd). Consider this m. Thus, $n = 2m + 1$, so that $n - 1 = 2m$. Thus, $n - 1$ is divisible by 2. In other words, $n \equiv 1 \bmod 2$ (by the definition of "$n \equiv 1 \bmod 2$"). This proves the implication (514).]

Next, we claim the following logical implication:

$$(n \equiv 1 \bmod 2) \implies (\text{the integer } n \text{ is odd}) . \tag{515}$$

[*Proof of (515):* Assume that $n \equiv 1 \bmod 2$. We shall show that the integer n is odd.

We have $n \equiv 1 \bmod 2$. In other words, $n - 1$ is divisible by 2 (by the definition of "$n \equiv 1 \bmod 2$"). In other words, there exists a $z \in \mathbb{Z}$ such that $n - 1 = 2z$. Consider this z. Now, $n - 1 = 2z$, so that $n = 2z + 1$. Hence, n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$ (namely, for $m = z$). Hence, the integer n is odd (since the integer n is odd if and only if n can be written in the form $n = 2m + 1$ for some $m \in \mathbb{Z}$). This proves the implication (515).]

Combining the two implications (514) and (515), we obtain the equivalence

$$(\text{the integer } n \text{ is odd}) \iff (n \equiv 1 \operatorname{mod} 2).$$

In other words, the integer n is odd if and only if $n \equiv 1 \operatorname{mod} 2$. This proves Proposition 2.160 **(b)**. □

Proof of Proposition 2.159. Let $u\%2$ denote the remainder of the division of u by 2. Then, Corollary 2.155 **(a)** (applied to $N = 2$ and $n = u$) yields that $u\%2 \in \{0, 1, \ldots, 2-1\}$ and $u\%2 \equiv u \operatorname{mod} 2$. Thus, $u\%2 \in \{0, 1, \ldots, 2-1\} = \{0, 1\}$.

Let $v\%2$ denote the remainder of the division of v by 2. Then, Corollary 2.155 **(a)** (applied to $N = 2$ and $n = v$) yields that $v\%2 \in \{0, 1, \ldots, 2-1\}$ and $v\%2 \equiv v \operatorname{mod} 2$. Thus, $v\%2 \in \{0, 1, \ldots, 2-1\} = \{0, 1\}$.

We have either $u\%2 = 0$ or $u\%2 = 1$ (since $u\%2 \in \{0, 1\}$). Thus, we are in one of the following two cases:

Case 1: We have $u\%2 = 0$.

Case 2: We have $u\%2 = 1$.

Let us first consider Case 1. In this case, we have $u\%2 = 0$.

From $u\%2 \equiv u \operatorname{mod} 2$, we obtain $u \equiv u\%2 = 0 \operatorname{mod} 2$. But Proposition 2.160 **(a)** (applied to $n = u$) shows that the integer u is even if and only if $u \equiv 0 \operatorname{mod} 2$. Thus, u is even (since $u \equiv 0 \operatorname{mod} 2$). Hence, Corollary 2.158 **(a)** (applied to $n = u$) yields $(-1)^u = 1$.

We have either $v\%2 = 0$ or $v\%2 = 1$ (since $v\%2 \in \{0, 1\}$). Thus, we are in one of the following two subcases:

Subcase 1.1: We have $v\%2 = 0$.

Subcase 1.2: We have $v\%2 = 1$.

Let us first consider Subcase 1.1. In this case, we have $v\%2 = 0$. But $u \equiv 0 = v\%2 \equiv v \operatorname{mod} 2$. Thus, the statement $(u \equiv v \operatorname{mod} 2)$ is true. Also, $v \equiv v\%2 = 0 \operatorname{mod} 2$.

Proposition 2.160 **(a)** (applied to $n = v$) shows that the integer v is even if and only if $v \equiv 0 \operatorname{mod} 2$. Thus, v is even (since $v \equiv 0 \operatorname{mod} 2$). Hence, Corollary 2.158 **(a)** (applied to $n = v$) yields $(-1)^v = 1$. Thus, $(-1)^u = 1 = (-1)^v$. Hence, the statement $\left((-1)^u = (-1)^v\right)$ is true.

Finally, the numbers u and v are either both even or both odd (since u and v are both even). Thus, the statement

$$(u \text{ and } v \text{ are either both even or both odd})$$

is true.

We have now shown that the three statements

$$(u \equiv v \operatorname{mod} 2), \ (u \text{ and } v \text{ are either both even or both odd}) \text{ and } \left((-1)^u = (-1)^v\right)$$

are all true. Thus, these three statements are equivalent. In other words, we have

$$\begin{aligned}(u \equiv v \bmod 2) &\iff (u \text{ and } v \text{ are either both even or both odd}) \\ &\iff \left((-1)^u = (-1)^v\right).\end{aligned}$$

Hence, Proposition 2.159 is proven in Subcase 1.1.

Let us next consider Subcase 1.2. In this case, we have $v\%2 = 1$. From $v\%2 \equiv v \bmod 2$, we obtain $v \equiv v\%2 = 1 \bmod 2$. From $u \equiv 0 \bmod 2$, we obtain $0 \equiv u \bmod 2$. Hence, if we had $u \equiv v \bmod 2$, then we would have $0 \equiv u \equiv v \equiv 1 \bmod 2$, which would contradict the fact that $0 \not\equiv 1 \bmod 2$ (since $2 \nmid 0 - 1$). Thus, we cannot have $u \equiv v \bmod 2$. In other words, the statement $(u \equiv v \bmod 2)$ is false.

Proposition 2.160 **(b)** (applied to $n = v$) shows that the integer v is odd if and only if $v \equiv 1 \bmod 2$. Thus, v is odd (since $v \equiv 1 \bmod 2$). Hence, Corollary 2.158 **(b)** (applied to $n = v$) yields $(-1)^v = -1$. Thus, $(-1)^u = 1 \neq -1 = (-1)^v$. Hence, the statement $\left((-1)^u = (-1)^v\right)$ is false.

Finally, the number v is not even (since v is odd), while the number u is not odd (since u is even). Hence, the numbers u and v are neither both even nor both odd. Thus, the statement

$$(u \text{ and } v \text{ are either both even or both odd})$$

is false.

We have now shown that the three statements

$$(u \equiv v \bmod 2), \ (u \text{ and } v \text{ are either both even or both odd}) \ \text{ and } \ \left((-1)^u = (-1)^v\right)$$

are all false. Thus, these three statements are equivalent. In other words, we have

$$\begin{aligned}(u \equiv v \bmod 2) &\iff (u \text{ and } v \text{ are either both even or both odd}) \\ &\iff \left((-1)^u = (-1)^v\right).\end{aligned}$$

Hence, Proposition 2.159 is proven in Subcase 1.2.

We have now proven Proposition 2.159 in each of the two Subcases 1.1 and 1.2. Since these two Subcases cover the whole Case 1, we thus conclude that Proposition 2.159 holds in Case 1.

The arguments needed to deal with Case 2 are very similar: We again must split the case into two subcases, the first being $v\%2 = 1$ and the second being $v\%2 = 0$; the reasoning in each subcase is completely analogous to the reasoning in the corresponding subcase of Case 1 (but with the roles of "even" and "odd" interchanged). We leave the details to the reader.

We have now proven Proposition 2.159 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Proposition 2.159 always holds. □

Solution to Exercise 2.4. We have proven Proposition 2.159 and Proposition 2.160; thus, Exercise 2.4 is solved. □

7.6. Solution to Exercise 2.5

Solution to Exercise 2.5. Let $n = h - p - 1$. Thus, $n \in \mathbb{Z}$. Let $n\%N$ denote the remainder of the division of n by N. Thus, Corollary 2.155 **(a)** yields that $n\%N \in \{0, 1, \ldots, N-1\}$ and $n\%N \equiv n \mod N$.

Let $x = (p+1) + (n\%N)$. Then, $x \in \mathbb{Z}$ (since both $p+1$ and $n\%N$ belong to $\mathbb{Z}$) and

$$\begin{aligned} x &= (p+1) + (n\%N) \\ &\in \{(p+1)+0, (p+1)+1, \ldots, (p+1)+(N-1)\} \\ &\qquad \left(\text{since } n\%N \in \{0,1,\ldots,N-1\}\right) \\ &= \{p+1, p+2, \ldots, p+N\}. \end{aligned}$$

Moreover, $x = (p+1) + \underbrace{(n\%N)}_{\equiv n = h-p-1 \mod N} \equiv (p+1) + (h-p-1) = h \mod N$. Hence, x is an element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$ (since $x \in \{p+1, p+2, \ldots, p+N\}$ and $x \equiv h \mod N$). Thus, there exists **at least one** element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$ (namely, $g = x$).

Now, let $g \in \{p+1, p+2, \ldots, p+N\}$ be such that $g \equiv h \mod N$. We shall prove that $g = x$.

Indeed, $\underbrace{g}_{\equiv h \mod N} - (p+1) \equiv h - (p+1) = h - p - 1 = n \mod N$. Also, from $g \in \{p+1, p+2, \ldots, p+N\}$, we obtain $g - (p+1) \in \{0, 1, \ldots, N-1\}$. Thus, Corollary 2.155 **(c)** (applied to $c = g - (p+1)$) yields $g - (p+1) = n\%N$. Hence, $g = (p+1) + (n\%N) = x$.

Now, forget that we fixed g. We thus have shown that if $g \in \{p+1, p+2, \ldots, p+N\}$ satisfies $g \equiv h \mod N$, then $g = x$. In other words, every element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$ must be equal to x. Thus, there exists **at most one** element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$.

We have now shown the following two facts:

- There exists **at least one** element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$.
- There exists **at most one** element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$.

Combining these two facts, we conclude that there exists a **unique** element $g \in \{p+1, p+2, \ldots, p+N\}$ satisfying $g \equiv h \mod N$. This solves Exercise 2.5. □

7.7. Solution to Exercise 2.6

Solution to Exercise 2.6. **(a)** We have $a - b = 1(a-b)$, thus $a - b \mid a - b$. In other words, $a \equiv b \mod a - b$ (by the definition of "congruent"). Hence, Proposition 2.22

(applied to $n = a - b$) yields $a^k \equiv b^k \mod a - b$. In other words, $a - b \mid a^k - b^k$ (by the definition of "congruent"). This solves Exercise 2.6 **(a)**.

(b) We know that k is odd. Hence, Corollary 2.158 **(b)** (applied to $n = k$) yields $(-1)^k = -1$.

Exercise 2.6 **(a)** (applied to $-b$ instead of b) yields $a - (-b) \mid a^k - (-b)^k$. In view of $a - (-b) = a + b$ and

$$a^k - \underbrace{(-b)^k}_{=(-1)^k b^k} = a^k - \underbrace{(-1)^k}_{=-1} b^k = a^k - (-1) b^k = a^k + b^k,$$

this rewrites as $a + b \mid a^k + b^k$. This solves Exercise 2.6 **(b)**. □

7.8. Solution to Exercise 2.7

Solution to Exercise 2.7. Proposition 2.66 **(a)** shows that we have

$$b_n \in \mathbb{N} \text{ for each } n \in \mathbb{N}. \qquad (516)$$

For every integer $m \geq 1$, we have

$$b_m^r + 1 = b_{m+1} b_{m-1}. \qquad (517)$$

(Indeed, this is precisely the identity (100), which was already proven during our proof of Proposition 2.66.)

Now, let n be a positive integer. Thus, the four integers $n-1, n, n+1, n+2$ all belong to $\mathbb{N}$. Hence, the numbers $b_{n-1}, b_n, b_{n+1}, b_{n+2}$ all belong to $\mathbb{N}$ (by (516)).

We have $n \geq 1$ (since n is a positive integer). Hence, (517) (applied to $m = n$) yields $b_n^r + 1 = b_{n+1} b_{n-1}$.

Define an integer x by $x = b_n + 1$. (This x is indeed an integer, because $b_n \in \mathbb{N} \subseteq \mathbb{Z}$.) We have[293] $b_n \equiv -1 \mod x$ (since $b_n - (-1) = b_n + 1 = x$ is divisible by x). Also, Exercise 2.6 **(b)** (applied to r, b_n and 1 instead of k, a and b) shows that $b_n + 1 \mid b_n^r + 1$. In view of $b_n + 1 = x$ and $b_n^r + 1 = b_{n+1} b_{n-1}$, this rewrites as $x \mid b_{n+1} b_{n-1}$. In other words, $b_{n+1} b_{n-1} \equiv 0 \mod x$. But $r \neq 0$ (since r is odd). Combining this with $r \geq 0$, we obtain $r > 0$. Hence, $r \geq 1$ (since r is an integer). Now,

$$b_{n-1} \underbrace{b_{n+1}^r}_{\substack{=b_{n+1} b_{n+1}^{r-1} \\ (\text{since } r \geq 1)}} = \underbrace{b_{n-1} b_{n+1}}_{\equiv 0 \mod x} b_{n+1}^{r-1} \equiv 0 \mod x. \qquad (518)$$

Now, $n + 1 \geq n \geq 1$. Thus, (517) (applied to $m = n + 1$) yields

$$b_{n+1}^r + 1 = \underbrace{b_{(n+1)+1}}_{=b_{n+2}} \underbrace{b_{(n+1)-1}}_{=b_n} = b_{n+2} b_n. \qquad (519)$$

[293] We recall that the four numbers $b_{n-1}, b_n, b_{n+1}, b_{n+2}$ are integers; this justifies the use of these four numbers in congruences modulo x.

We have

$$b_n b_{n-1} (b_{n+2}+1) = b_{n-1} \underbrace{b_{n+2}b_n}_{\substack{=b_{n+1}^r+1 \\ \text{(by (519))}}} + \underbrace{b_n}_{\equiv -1 \bmod x} b_{n-1}$$
$$\equiv b_{n-1}\left(b_{n+1}^r+1\right) + (-1) b_{n-1} = b_{n-1}b_{n+1}^r \equiv 0 \bmod x$$

(by (518)). But

$$\underbrace{x}_{=b_n+1} b_{n-1}(b_{n+2}+1) = (b_n+1) b_{n-1} (b_{n-2}+1)$$
$$= \underbrace{b_n b_{n-1}(b_{n+2}+1)}_{\equiv 0 \bmod x} + b_{n-1}(b_{n-2}+1) \equiv b_{n-1}(b_{n-2}+1) \bmod x.$$

Hence,

$$b_{n-1}(b_{n-2}+1) \equiv \underbrace{x}_{\equiv 0 \bmod x} b_{n-1}(b_{n+2}+1) \equiv 0 \bmod x.$$

In other words, $x \mid b_{n-1}(b_{n-2}+1)$. In view of $x = b_n + 1$, this rewrites as $b_n + 1 \mid b_{n-1}(b_{n+2}+1)$. This solves Exercise 2.7. □

7.9. Solution to Exercise 2.8

To prove Theorem 2.161, we just mildly adapt our proof of Theorem 2.53:

Proof of Theorem 2.161. For any $n \in \mathbb{N}$, we have $g - n \in \mathbb{Z}_{\leq g}$ [294]. Hence, for each $n \in \mathbb{N}$, we can define a logical statement $\mathcal{B}(n)$ by

$$\mathcal{B}(n) = \mathcal{A}(g-n).$$

Consider this $\mathcal{B}(n)$.

Now, let us consider the Assumptions A and B from Corollary 2.54. We claim that both of these assumptions are satisfied.

Indeed, the statement $\mathcal{A}(g)$ holds (by Assumption 1). But the definition of the statement $\mathcal{B}(0)$ shows that $\mathcal{B}(0) = \mathcal{A}(g-0) = \mathcal{A}(g)$. Hence, the statement $\mathcal{B}(0)$ holds (since the statement $\mathcal{A}(g)$ holds). In other words, Assumption A is satisfied.

Now, we shall show that Assumption B is satisfied. Indeed, let $p \in \mathbb{N}$ be such that $\mathcal{B}(p)$ holds. The definition of the statement $\mathcal{B}(p)$ shows that $\mathcal{B}(p) = \mathcal{A}(g-p)$. Hence, the statement $\mathcal{A}(g-p)$ holds (since $\mathcal{B}(p)$ holds).

Also, $p \in \mathbb{N}$, so that $p \geq 0$ and thus $g - p \leq g$. In other words, $g - p \in \mathbb{Z}_{\leq g}$ (since $\mathbb{Z}_{\leq g}$ is the set of all integers that are $\leq g$).

[294] *Proof.* Let $n \in \mathbb{N}$. Thus, $n \geq 0$, so that $g - \underbrace{n}_{\geq 0} \leq g - 0 = g$. Hence, $g - n$ is an integer $\leq g$. In other words, $g - n \in \mathbb{Z}_{\leq g}$ (since $\mathbb{Z}_{\leq g}$ is the set of all integers that are $\leq g$). Qed.

Recall that Assumption 2 holds. In other words, if $m \in \mathbb{Z}_{\leq g}$ is such that $\mathcal{A}(m)$ holds, then $\mathcal{A}(m-1)$ also holds. Applying this to $m = g - p$, we conclude that $\mathcal{A}((g-p)-1)$ holds (since $\mathcal{A}(g-p)$ holds).

But the definition of $\mathcal{B}(p+1)$ yields $\mathcal{B}(p+1) = \mathcal{A}\left(\underbrace{g-(p+1)}_{=(g-p)-1}\right) = \mathcal{A}((g-p)-1)$. Hence, the statement $\mathcal{B}(p+1)$ holds (since the statement $\mathcal{A}((g-p)-1)$ holds).

Now, forget that we fixed p. We thus have shown that if $p \in \mathbb{N}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p+1)$ also holds. In other words, Assumption B is satisfied.

We now know that both Assumption A and Assumption B are satisfied. Hence, Corollary 2.54 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{N}. \tag{520}$$

Now, let $n \in \mathbb{Z}_{\leq g}$. Thus, n is an integer such that $n \leq g$ (by the definition of $\mathbb{Z}_{\leq g}$). Hence, $g - n \geq 0$, so that $g - n \in \mathbb{N}$. Thus, (520) (applied to $g - n$ instead of n) yields that $\mathcal{B}(g-n)$ holds. But the definition of $\mathcal{B}(g-n)$ yields $\mathcal{B}(g-n) = \mathcal{A}\left(\underbrace{g-(g-n)}_{=n}\right) = \mathcal{A}(n)$. Hence, the statement $\mathcal{A}(n)$ holds (since $\mathcal{B}(g-n)$ holds).

Now, forget that we fixed n. We thus have shown that $\mathcal{A}(n)$ holds for each $n \in \mathbb{Z}_{\leq g}$. This proves Theorem 2.161. □

Let us restate Theorem 2.161 as follows:

Corollary 7.5. Let $h \in \mathbb{Z}$. Let $\mathbb{Z}_{\leq h}$ be the set $\{h, h-1, h-2, \ldots\}$ (that is, the set of all integers that are $\leq h$). For each $n \in \mathbb{Z}_{\geq h}$, let $\mathcal{B}(n)$ be a logical statement.

Assume the following:

Assumption A: The statement $\mathcal{B}(h)$ holds.

Assumption B: If $p \in \mathbb{Z}_{\leq h}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p-1)$ also holds.

Then, $\mathcal{B}(n)$ holds for each $n \in \mathbb{Z}_{\leq h}$.

Proof of Corollary 7.5. Corollary 7.5 is exactly Theorem 2.161, except that some names have been changed:

- The variable g has been renamed as h.
- The statements $\mathcal{A}(n)$ have been renamed as $\mathcal{B}(n)$.
- Assumption 1 and Assumption 2 have been renamed as Assumption A and Assumption B.

- The variable m in Assumption B has been renamed as p.

Thus, Corollary 7.5 holds (since Theorem 2.161 holds). □

In order to prove Theorem 2.162, we modify our proof of Theorem 2.74 as follows:

Proof of Theorem 2.162. Let $\mathbb{Z}_{\leq h}$ be the set $\{h, h-1, h-2, \ldots\}$ (that is, the set of all integers that are $\leq h$).

For each $n \in \mathbb{Z}_{\leq h}$, we define $\mathcal{B}(n)$ to be the logical statement

$$(\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds}).$$

Now, let us consider the Assumptions A and B from Corollary 7.5. We claim that both of these assumptions are satisfied.

Assumption 1 says that if $g \leq h$, then the statement $\mathcal{A}(h)$ holds. Thus, $\mathcal{B}(h)$ holds[295]. In other words, Assumption A is satisfied.

Next, we shall prove that Assumption B is satisfied. Indeed, let $p \in \mathbb{Z}_{\leq h}$ be such that $\mathcal{B}(p)$ holds. We shall now show that $\mathcal{B}(p-1)$ also holds.

Indeed, assume that $p-1 \in \{g, g+1, \ldots, h\}$. Thus, $p-1 \geq g$, so that $p \geq p-1 \geq g$. Combining this with $p \leq h$ (since $p \in \mathbb{Z}_{\leq h}$), we conclude that $p \in \{g, g+1, \ldots, h\}$ (since p is an integer). But we have assumed that $\mathcal{B}(p)$ holds. In other words,

$$\text{if } p \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(p) \text{ holds}$$

(because the statement $\mathcal{B}(p)$ is defined as $(\text{if } p \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(p) \text{ holds})$). Thus, $\mathcal{A}(p)$ holds (since we have $p \in \{g, g+1, \ldots, h\}$). Also, from $p-1 \geq g$, we obtain $p \geq g+1$. Combining this with $p \leq h$, we find $p \in \{g+1, g+2, \ldots, h\}$. Thus, we know that $p \in \{g+1, g+2, \ldots, h\}$ is such that $\mathcal{A}(p)$ holds. Hence, Assumption 2 (applied to $m = p$) shows that $\mathcal{A}(p-1)$ also holds.

Now, forget that we assumed that $p-1 \in \{g, g+1, \ldots, h\}$. We thus have proven that if $p-1 \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(p-1)$ holds. In other words, $\mathcal{B}(p-1)$ holds (since the statement $\mathcal{B}(p-1)$ is defined as $(\text{if } p-1 \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(p-1) \text{ holds})$).

Now, forget that we fixed p. We thus have proven that if $p \in \mathbb{Z}_{\leq h}$ is such that $\mathcal{B}(p)$ holds, then $\mathcal{B}(p-1)$ also holds. In other words, Assumption B is satisfied.

We now know that both Assumption A and Assumption B are satisfied. Hence, Corollary 7.5 shows that

$$\mathcal{B}(n) \text{ holds for each } n \in \mathbb{Z}_{\leq h}. \tag{521}$$

[295] *Proof.* Assume that $h \in \{g, g+1, \ldots, h\}$. Thus, $g \leq h$. But Assumption 1 says that if $g \leq h$, then the statement $\mathcal{A}(h)$ holds. Hence, the statement $\mathcal{A}(h)$ holds (since $g \leq h$).

Now, forget that we assumed that $h \in \{g, g+1, \ldots, h\}$. We thus have proven that if $h \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(h)$ holds. In other words, $\mathcal{B}(h)$ holds (because the statement $\mathcal{B}(h)$ is defined as $(\text{if } h \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(h) \text{ holds})$). Qed.

Now, let $n \in \{g, g+1, \ldots, h\}$. Thus, $n \leq h$, so that $n \in \mathbb{Z}_{\leq h}$. Hence, (521) shows that $\mathcal{B}(n)$ holds. In other words,

$$\text{if } n \in \{g, g+1, \ldots, h\}, \text{ then } \mathcal{A}(n) \text{ holds}$$

(since the statement $\mathcal{B}(n)$ was defined as (if $n \in \{g, g+1, \ldots, h\}$, then $\mathcal{A}(n)$ holds)). Thus, $\mathcal{A}(n)$ holds (since we have $n \in \{g, g+1, \ldots, h\}$).

Now, forget that we fixed n. We thus have shown that $\mathcal{A}(n)$ holds for each $n \in \{g, g+1, \ldots, h\}$. This proves Theorem 2.162. □

Solution to Exercise 2.8. Theorem 2.161 and Theorem 2.162 have been proven. Thus, Exercise 2.8 is solved. □

7.10. Solution to Exercise 2.9

First solution to Exercise 2.9. We shall solve Exercise 2.9 by induction on n:

Induction base: Comparing

$$\begin{cases} 0/2+1, & \text{if } 0 \text{ is even;} \\ -(0+1)/2, & \text{if } 0 \text{ is odd} \end{cases} = 0/2+1 \qquad (\text{since } 0 \text{ is even})$$
$$= 1$$

with $\sum_{k=0}^{0} (-1)^k (k+1) = \underbrace{(-1)^0}_{=1} \underbrace{(0+1)}_{=1} = 1$, we find

$$\sum_{k=0}^{0} (-1)^k (k+1) = \begin{cases} 0/2+1, & \text{if } 0 \text{ is even;} \\ -(0+1)/2, & \text{if } 0 \text{ is odd} \end{cases}.$$

In other words, Exercise 2.9 holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{Z}_{\geq 1}$. Assume that Exercise 2.9 holds for $n = m-1$. We must prove that Exercise 2.9 holds for $n = m$.

We have assumed that Exercise 2.9 holds for $n = m-1$. In other words,

$$\sum_{k=0}^{m-1} (-1)^k (k+1) = \begin{cases} (m-1)/2+1, & \text{if } m-1 \text{ is even;} \\ -((m-1)+1)/2, & \text{if } m-1 \text{ is odd} \end{cases}. \tag{522}$$

We must prove that Exercise 2.9 holds for $n = m$. In other words, we must prove that

$$\sum_{k=0}^{m} (-1)^k (k+1) = \begin{cases} m/2+1, & \text{if } m \text{ is even;} \\ -(m+1)/2, & \text{if } m \text{ is odd} \end{cases}. \tag{523}$$

The integer m is either even or odd. Thus, we are in one of the following two cases:

Case 1: The integer m is even.

Case 2: The integer m is odd.

Let us first consider Case 1. In this case, the integer m is even. Thus, the integer $m-1$ is odd[296]. Hence, (522) becomes

$$\sum_{k=0}^{m-1} (-1)^k (k+1) = \begin{cases} (m-1)/2+1, & \text{if } m-1 \text{ is even;} \\ -((m-1)+1)/2, & \text{if } m-1 \text{ is odd} \end{cases}$$
$$= -((m-1)+1)/2 \qquad (\text{since } m-1 \text{ is odd}).$$

Also, Corollary 2.158 **(a)** (applied to $n=m$) yields $(-1)^m = 1$ (since m is even).

Now, we can split off the addend for $k=m$ from the sum $\sum_{k=0}^{m} (-1)^k (k+1)$ (since $m \in \{0,1,\ldots,m\}$). We thus obtain

$$\sum_{k=0}^{m} (-1)^k (k+1) = \underbrace{\sum_{k=0}^{m-1} (-1)^k (k+1)}_{=-((m-1)+1)/2} + \underbrace{(-1)^m}_{=1} (m+1)$$
$$= -((m-1)+1)/2 + (m+1) = m/2+1.$$

Comparing this with

$$\begin{cases} m/2+1, & \text{if } m \text{ is even;} \\ -(m+1)/2, & \text{if } m \text{ is odd} \end{cases} = m/2+1 \qquad (\text{since } m \text{ is even}),$$

we obtain

$$\sum_{k=0}^{m} (-1)^k (k+1) = \begin{cases} m/2+1, & \text{if } m \text{ is even;} \\ -(m+1)/2, & \text{if } m \text{ is odd} \end{cases}.$$

Thus, (523) holds in Case 1.

Let us next consider Case 1. In this case, the integer m is odd. Thus, the integer $m-1$ is even[297]. Hence, (522) becomes

$$\sum_{k=0}^{m-1} (-1)^k (k+1) = \begin{cases} (m-1)/2+1, & \text{if } m-1 \text{ is even;} \\ -((m-1)+1)/2, & \text{if } m-1 \text{ is odd} \end{cases}$$
$$= (m-1)/2+1 \qquad (\text{since } m-1 \text{ is even}).$$

[296] *Proof.* Proposition 2.160 **(a)** (applied to $n=m$) shows that the integer m is even if and only if $m \equiv 0 \mod 2$. Hence, $m \equiv 0 \mod 2$ (since the integer m is even). Thus, $\underbrace{m}_{\equiv 0 \mod 2} - 1 \equiv 0 - 1 = -1 \equiv 1 \mod 2$ (since $2 \mid (-1) - 1$).

But Proposition 2.160 **(b)** (applied to $n = m-1$) shows that the integer $m-1$ is odd if and only if $m-1 \equiv 1 \mod 2$. Hence, the integer $m-1$ is odd (since $m-1 \equiv 1 \mod 2$).

[297] *Proof.* Proposition 2.160 **(b)** (applied to $n=m$) shows that the integer m is odd if and only if $m \equiv 1 \mod 2$. Hence, $m \equiv 1 \mod 2$ (since the integer m is odd). Thus, $\underbrace{m}_{\equiv 1 \mod 2} - 1 \equiv 1 - 1 = 0 \mod 2$.

But Proposition 2.160 **(a)** (applied to $n = m-1$) shows that the integer $m-1$ is even if and only if $m-1 \equiv 0 \mod 2$. Hence, the integer $m-1$ is even (since $m-1 \equiv 0 \mod 2$).

Also, Corollary 2.158 **(b)** (applied to $n = m$) yields $(-1)^m = -1$ (since m is odd).

Now, we can split off the addend for $k = m$ from the sum $\sum_{k=0}^{m} (-1)^k (k+1)$ (since $m \in \{0, 1, \ldots, m\}$). We thus obtain

$$\sum_{k=0}^{m} (-1)^k (k+1) = \underbrace{\sum_{k=0}^{m-1} (-1)^k (k+1)}_{=(m-1)/2+1} + \underbrace{(-1)^m}_{=-1} (m+1)$$
$$= (m-1)/2 + 1 + (-1)(m+1) = -(m+1)/2.$$

Comparing this with

$$\begin{cases} m/2+1, & \text{if } m \text{ is even;} \\ -(m+1)/2, & \text{if } m \text{ is odd} \end{cases} = -(m+1)/2 \qquad (\text{since } m \text{ is odd}),$$

we obtain

$$\sum_{k=0}^{m} (-1)^k (k+1) = \begin{cases} m/2+1, & \text{if } m \text{ is even;} \\ -(m+1)/2, & \text{if } m \text{ is odd} \end{cases}.$$

Thus, (523) holds in Case 2.

We thus have proven that (523) holds in each of the two Cases 1 and 2. Thus, (523) always holds (since Cases 1 and 2 cover all possibilities). In other words, Exercise 2.9 holds for $n = m$. This completes the induction step. Thus, Exercise 2.9 is proven by induction. □

Second solution to Exercise 2.9. We have

$$\sum_{k=0}^{n} (-1)^k (k+1)$$
$$= (-1)^0 1 + (-1)^1 2 + (-1)^2 3 + (-1)^3 4 + \cdots + (-1)^n (n+1)$$
$$= 1 - 2 + 3 - 4 \pm \cdots + (-1)^n (n+1). \tag{524}$$

Notice that the sign of the last addend on the right hand side depends on whether n is even or odd. Thus, we distinguish between the following two cases:

Case 1: The integer n is even.

Case 2: The integer n is odd.

Let us first consider Case 1. In this case, the integer n is even. Thus, $(-1)^n = 1$

(by Corollary 2.158 **(a)**). Hence, (524) becomes

$$\begin{aligned}
\sum_{k=0}^{n} (-1)^k (k+1) &= 1-2+3-4 \pm \cdots + \underbrace{(-1)^n}_{=1} (n+1) \\
&= 1-2+3-4 \pm \cdots + (n+1) \\
&= \underbrace{(1-2)}_{=-1} + \underbrace{(3-4)}_{=-1} + \underbrace{(5-6)}_{=-1} + \cdots + \underbrace{((n-1)-n)}_{=-1} + (n+1) \\
&= \underbrace{\underbrace{((-1)+(-1)+(-1)+\cdots+(-1))}_{n/2 \text{ addends}}}_{=n/2\cdot(-1)=-n/2} + (n+1) \\
&= -n/2 + (n+1) = n/2+1.
\end{aligned}$$

Comparing this with

$$\begin{cases} n/2+1, & \text{if } n \text{ is even;} \\ -(n+1)/2, & \text{if } n \text{ is odd} \end{cases} = n/2+1 \qquad (\text{since } n \text{ is even}),$$

we obtain $\sum_{k=0}^{n} (-1)^k (k+1) = \begin{cases} n/2+1, & \text{if } n \text{ is even;} \\ -(n+1)/2, & \text{if } n \text{ is odd} \end{cases}$. Hence, Exercise 2.9 is solved in Case 1.

Let us next consider Case 2. In this case, the integer n is odd. Thus, $(-1)^n = -1$ (by Corollary 2.158 **(b)**). Hence, (524) becomes

$$\begin{aligned}
\sum_{k=0}^{n} (-1)^k (k+1) &= 1-2+3-4 \pm \cdots + \underbrace{(-1)^n}_{=-1} (n+1) \\
&= 1-2+3-4 \pm \cdots - (n+1) \\
&= \underbrace{(1-2)}_{=-1} + \underbrace{(3-4)}_{=-1} + \underbrace{(5-6)}_{=-1} + \cdots + \underbrace{(n-(n+1))}_{=-1} \\
&= \underbrace{((-1)+(-1)+(-1)+\cdots+(-1))}_{(n+1)/2 \text{ addends}} \\
&= (n+1)/2 \cdot (-1) = -(n+1)/2.
\end{aligned}$$

Comparing this with

$$\begin{cases} n/2+1, & \text{if } n \text{ is even;} \\ -(n+1)/2, & \text{if } n \text{ is odd} \end{cases} = -(n+1)/2 \qquad (\text{since } n \text{ is odd}),$$

we obtain $\sum_{k=0}^{n} (-1)^k (k+1) = \begin{cases} n/2+1, & \text{if } n \text{ is even;} \\ -(n+1)/2, & \text{if } n \text{ is odd} \end{cases}$. Hence, Exercise 2.9 is solved in Case 2.

We thus have solved Exercise 2.9 in both Cases 1 and 2. Hence, Exercise 2.9 always holds. □

7.11. Solution to Exercise 3.1

Solution to Exercise 3.1. For each $i \in \{0, 1, \ldots, m\}$, define an integer $s_i \in \mathbb{N}$ by $s_i = k_1 + k_2 + \cdots + k_i$. We shall prove that

$$\frac{(k_1 + k_2 + \cdots + k_m)!}{k_1! k_2! \cdots k_m!} = \prod_{i=1}^{m} \binom{s_i}{k_i}. \tag{525}$$

[*Proof of (525):* If $m = 0$, then (525) holds[298]. Hence, for the rest of the proof of (525), we WLOG assume that we don't have $m = 0$.

We have $m \geq 1$ (since $m \in \mathbb{N}$ but we don't have $m = 0$), so that $m - 1 \geq 0$.

The definition of s_0 yields $s_0 = k_1 + k_2 + \cdots + k_0 = (\text{empty sum}) = 0$. Hence, $s_0! = 0! = 1$. Now,

$$\begin{aligned}
\prod_{i=1}^{m} \frac{1}{s_{i-1}!} &= \prod_{i=0}^{m-1} \frac{1}{s_i!} \qquad (\text{here, we have substituted } i \text{ for } i-1 \text{ in the product}) \\
&= \underbrace{\frac{1}{s_0!}}_{\substack{=1 \\ (\text{since } s_0! = 1)}} \prod_{i=1}^{m-1} \frac{1}{s_i!} \qquad (\text{since } m - 1 \geq 0) \\
&= \prod_{i=1}^{m-1} \frac{1}{s_i!} = \frac{1}{\prod_{i=1}^{m-1} s_i!}.
\end{aligned} \tag{526}$$

The definition of s_m yields $s_m = k_1 + k_2 + \cdots + k_m$.

Each $i \in \{1, 2, \ldots, m\}$ satisfies

$$\binom{s_i}{k_i} = s_i! \cdot \frac{1}{k_i!} \cdot \frac{1}{s_{i-1}!}$$

[298] *Proof.* This is just an exercise in understanding empty sums and empty products: Empty sums such as $k_1 + k_2 + \cdots + k_0$ are defined to be 0, and empty products such as $k_1! k_2! \cdots k_0!$ or $\prod_{i=1}^{0} \binom{s_i}{k_i}$ are defined to be 1. With this in mind (and remembering that $0! = 1$), it becomes completely straightforward to verify (525) when $m = 0$.

[299]. Hence,

$$\prod_{i=1}^{m} \underbrace{\binom{s_i}{k_i}}_{=s_i!\cdot\frac{1}{k_i!}\cdot\frac{1}{s_{i-1}!}} = \prod_{i=1}^{m}\left(s_i!\cdot\frac{1}{k_i!}\cdot\frac{1}{s_{i-1}!}\right) = \underbrace{\left(\prod_{i=1}^{m} s_i!\right)}_{\substack{=\left(\prod_{i=1}^{m-1} s_i!\right)s_m! \\ (\text{since } m\geq 1)}} \cdot \left(\prod_{i=1}^{m}\frac{1}{k_i!}\right)\cdot \underbrace{\left(\prod_{i=1}^{m}\frac{1}{s_{i-1}!}\right)}_{\substack{=\frac{1}{\prod_{i=1}^{m-1} s_i!} \\ (\text{by (526)})}}$$

$$= \left(\prod_{i=1}^{m-1} s_i!\right) s_m!\cdot\left(\prod_{i=1}^{m}\frac{1}{k_i!}\right)\cdot\frac{1}{\prod_{i=1}^{m-1} s_i!}$$

$$= \underbrace{s_m}_{=k_1+k_2+\cdots+k_m}!\cdot\underbrace{\left(\prod_{i=1}^{m}\frac{1}{k_i!}\right)}_{=\frac{1}{\prod_{i=1}^{m} k_i!}=\frac{1}{k_1!k_2!\cdots k_m!}}$$

$$= (k_1+k_2+\cdots+k_m)!\cdot\frac{1}{k_1!k_2!\cdots k_m!} = \frac{(k_1+k_2+\cdots+k_m)!}{k_1!k_2!\cdots k_m!}.$$

This proves (525).]

On the other hand, each $i\in\{1,2,\ldots,m\}$ satisfies $\binom{s_i}{k_i}\in\mathbb{Z}$ (by Proposition 3.20 (applied to s_i and k_i instead of m and n)). In other words, for each $i\in\{1,2,\ldots,m\}$, the number $\binom{s_i}{k_i}$ is an integer. Hence, $\prod_{i=1}^{m}\binom{s_i}{k_i}$ is an integer (since a product of integers always is an integer). In other words, $\frac{(k_1+k_2+\cdots+k_m)!}{k_1!k_2!\cdots k_m!}$ is an integer

[299] *Proof:* Let $i\in\{1,2,\ldots,m\}$. Then, both $i-1$ and i belong to the set $\{0,1,\ldots,m\}$. Hence, the definition of s_{i-1} yields $s_{i-1}=k_1+k_2+\cdots+k_{i-1}\in\mathbb{N}$ (since $k_1,k_2,\ldots,k_{i-1}$ are elements of $\mathbb{N}$). Thus, $s_{i-1}\geq 0$. Meanwhile, the definition of s_i yields

$$s_i = k_1+k_2+\cdots+k_i = \underbrace{(k_1+k_2+\cdots+k_{i-1})}_{=s_{i-1}}+k_i = \underbrace{s_{i-1}}_{\geq 0}+k_i \geq k_i.$$

Also, $s_i=k_1+k_2+\cdots+k_i\in\mathbb{N}$ (since $k_1,k_2,\ldots,k_i$ are elements of $\mathbb{N}$). From $s_i=s_{i-1}+k_i$, we obtain $s_i-k_i=s_{i-1}$.

Now, (229) (applied to s_i and k_i instead of m and n) yields

$$\binom{s_i}{k_i} = \frac{s_i!}{k_i!\,(s_i-k_i)!} = \frac{s_i!}{k_i!s_{i-1}!} \qquad (\text{since } s_i-k_i=s_{i-1})$$
$$= s_i!\cdot\frac{1}{k_i!}\cdot\frac{1}{s_{i-1}!}.$$

(since $\dfrac{(k_1+k_2+\cdots+k_m)!}{k_1!k_2!\cdots k_m!} = \prod_{i=1}^{m} \dbinom{s_i}{k_i}$). Since $\dfrac{(k_1+k_2+\cdots+k_m)!}{k_1!k_2!\cdots k_m!}$ is clearly positive (because the numbers $(k_1+k_2+\cdots+k_m)!$, $k_1!$, $k_2!$, $\ldots$, $k_m!$ are all positive), we can thus conclude that $\dfrac{(k_1+k_2+\cdots+k_m)!}{k_1!k_2!\cdots k_m!}$ is a positive integer. This solves Exercise 3.1. □

7.12. Solution to Exercise 3.2

7.12.1. The solution

We are going to give two proofs for each part of Exercise 3.2: one by direct manipulation of products, and one by induction on n. The induction proofs will rely on the following fact:

Lemma 7.6. Let m be a positive integer. Let $q \in \mathbb{Q}$. Then,

$$\binom{q}{m} = \frac{q-m+1}{m}\binom{q}{m-1}.$$

Proof of Lemma 7.6. We have $m-1 \in \mathbb{N}$ (since m is a positive integer). Thus, (226) (applied to q and $m-1$ instead of m and n) yields

$$\binom{q}{m-1} = \frac{q(q-1)\cdots(q-(m-1)+1)}{(m-1)!} = \frac{q(q-1)\cdots(q-m+2)}{(m-1)!} \qquad (527)$$

(since $q-(m-1)+1 = q-m+2$).

On the other hand, (37) (applied to $n=m$) yields $m! = m\cdot(m-1)!$. Also, (226) (applied to q and m instead of m and n) yields

$$\binom{q}{m} = \frac{q(q-1)\cdots(q-m+1)}{m!} = \underbrace{\frac{1}{m!}}_{\substack{=\dfrac{1}{m\cdot(m-1)!} \\ (\text{since } m!=m\cdot(m-1)!)}} \underbrace{(q(q-1)\cdots(q-m+1))}_{\substack{=(q(q-1)\cdots(q-m+2))\cdot(q-m+1) \\ (\text{since } m \text{ is a positive integer})}}$$

$$= \frac{1}{m\cdot(m-1)!}(q(q-1)\cdots(q-m+2))\cdot(q-m+1)$$

$$= \frac{q-m+1}{m}\cdot\underbrace{\frac{q(q-1)\cdots(q-m+2)}{(m-1)!}}_{\substack{=\dbinom{q}{m-1} \\ (\text{by } (527))}} = \frac{q-m+1}{m}\binom{q}{m-1}.$$

This proves Lemma 7.6. □

We shall also use the following simple observation in the solutions to Exercise 3.2 **(b)**:

Lemma 7.7. Let $n \in \mathbb{N}$. Then,

$$\binom{2n}{n} = \frac{(2n)!}{n!^2}.$$

Proof of Lemma 7.7. From $n \in \mathbb{N}$, we obtain $n \geq 0$, and thus $2n \geq n \geq 0$. Hence, $2n \in \mathbb{N}$. Thus, Proposition 3.4 (applied to $m = 2n$) yields

$$\binom{2n}{n} = \frac{(2n)!}{n!\,(2n-n)!} = \frac{(2n)!}{n!n!}$$

(since $2n - n = n$). Thus, $\binom{2n}{n} = \frac{(2n)!}{n!n!} = \frac{(2n)!}{n!^2}$. This proves Lemma 7.7. $\square$

Solution to Exercise 3.2. **(a)** *First solution to Exercise 3.2* ***(a)****:* We have

$$\begin{aligned}
(2n)! &= 1 \cdot 2 \cdot \cdots \cdot (2n) = \prod_{k \in \{1,2,\ldots,2n\}} k = \underbrace{\left(\prod_{\substack{k \in \{1,2,\ldots,2n\}; \\ k \text{ is even}}} k \right)}_{\substack{=2 \cdot 4 \cdot 6 \cdot \cdots \cdot (2n) \\ = \prod_{i=1}^{n} (2i) \\ = 2^n \prod_{i=1}^{n} i}} \underbrace{\left(\prod_{\substack{k \in \{1,2,\ldots,2n\}; \\ k \text{ is odd}}} k \right)}_{=1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1)} \\
&\qquad \left(\begin{array}{c} \text{here, we have split the product } \prod_{k \in \{1,2,\ldots,2n\}} k \text{ into one product} \\ \text{containing all even } k \text{ and one product containing all odd } k \end{array} \right) \\
&= 2^n \underbrace{\left(\prod_{i=1}^{n} i \right)}_{=1 \cdot 2 \cdot \cdots \cdot n = n!} \cdot (1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1)) \\
&= 2^n n! \cdot (1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1)).
\end{aligned}$$

Dividing this equality by $2^n n!$, we obtain

$$\frac{(2n)!}{2^n n!} = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1) = (2n-1) \cdot (2n-3) \cdot \cdots \cdot 1.$$

This solves Exercise 3.2 **(a)**.

Second solution to Exercise 3.2 ***(a)****:* Let us solve Exercise 3.2 **(a)** by induction on n:

Induction base: From $2 \cdot 0 = 0$, we obtain

$$\frac{(2 \cdot 0)!}{2^0 0!} = \frac{0!}{2^0 0!} = \frac{1}{2^0} = \frac{1}{1} \qquad \left(\text{since } 2^0 = 1\right)$$
$$= 1.$$

Comparing this with

$$(2 \cdot 0 - 1) \cdot (2 \cdot 0 - 3) \cdot \cdots \cdot 1 = (\text{empty product}) = 1,$$

we find $(2 \cdot 0 - 1) \cdot (2 \cdot 0 - 3) \cdot \cdots \cdot 1 = \frac{(2 \cdot 0)!}{2^0 0!}$. In other words, Exercise 3.2 **(a)** holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Exercise 3.2 **(a)** holds for $n = m$. We must prove that Exercise 3.2 **(a)** holds for $n = m + 1$.

We have assumed that Exercise 3.2 **(a)** holds for $n = m$. In other words, we have

$$(2m - 1) \cdot (2m - 3) \cdot \cdots \cdot 1 = \frac{(2m)!}{2^m m!}. \tag{528}$$

Clearly, $m + 1$ is a positive integer (since $m \in \mathbb{N}$). Hence, (37) (applied to $n = m + 1$) yields $(m+1)! = (m+1) \cdot \left(\underbrace{(m+1) - 1}_{=m}\right)! = (m+1) \cdot m!$.

But $2(m+1) = 2m + 2$, and thus

$$\begin{aligned} (2(m+1))! &= (2m+2)! = 1 \cdot 2 \cdot \cdots \cdot (2m+2) \\ &= \underbrace{(1 \cdot 2 \cdot \cdots \cdot (2m))}_{\substack{=(2m)! \\ (\text{since } (2m)! = 1 \cdot 2 \cdot \cdots \cdot (2m))}} \cdot (2m+1) \cdot (2m+2) \\ &= (2m)! \cdot (2m+1) \cdot (2m+2). \end{aligned}$$

Now,

$$\begin{aligned} \frac{(2(m+1))!}{2^{m+1}(m+1)!} &= \underbrace{(2(m+1))!}_{=(2m)! \cdot (2m+1) \cdot (2m+2)} \Big/ \left(\underbrace{2^{m+1}}_{=2 \cdot 2^m} \underbrace{(m+1)!}_{=(m+1) \cdot m!}\right) \\ &= (2m)! \cdot (2m+1) \cdot (2m+2) \,/\, (2 \cdot 2^m \cdot (m+1) \cdot m!) \\ &= \frac{(2m)! \cdot (2m+1) \cdot (2m+2)}{2 \cdot 2^m \cdot (m+1) \cdot m!} = \frac{(2m)!}{2^m m!} \cdot (2m+1) \cdot \underbrace{\frac{2m+2}{2(m+1)}}_{=1} \\ &= \frac{(2m)!}{2^m m!} \cdot (2m+1) = (2m+1) \cdot \frac{(2m)!}{2^m m!}. \end{aligned}$$

Comparing this with

$$\begin{aligned}
&(2(m+1)-1)\cdot(2(m+1)-3)\cdot\cdots\cdot 1 \\
&= \underbrace{(2(m+1)-1)}_{=2m+1}\cdot\underbrace{((2(m+1)-3)\cdot(2(m+1)-5)\cdot\cdots\cdot 1)}_{\substack{=(2m-1)\cdot(2m-3)\cdot\cdots\cdot 1\\ (\text{since } 2(m+1)-3=2m-1)}} \\
&= (2m+1)\cdot\underbrace{((2m-1)\cdot(2m-3)\cdot\cdots\cdot 1)}_{\substack{=\frac{(2m)!}{2^m m!}\\ (\text{by (528)})}} = (2m+1)\cdot\frac{(2m)!}{2^m m!},
\end{aligned}$$

we obtain

$$(2(m+1)-1)\cdot(2(m+1)-3)\cdot\cdots\cdot 1 = \frac{(2(m+1))!}{2^{m+1}(m+1)!}.$$

In other words, Exercise 3.2 **(a)** holds for $n=m+1$. This completes the induction step. Thus, Exercise 3.2 **(a)** is solved.

(b) *First solution to Exercise 3.2 **(b)**:* The equality (226) (applied to $-1/2$ instead of m) yields

$$\begin{aligned}
\binom{-1/2}{n} &= \frac{(-1/2)(-1/2-1)\cdots(-1/2-n+1)}{n!} \\
&= \frac{1}{n!}\cdot\underbrace{(-1/2)(-1/2-1)\cdots(-1/2-n+1)}_{=\prod_{k=0}^{n-1}(-1/2-k)} \\
&= \frac{1}{n!}\cdot\prod_{k=0}^{n-1}\underbrace{(-1/2-k)}_{=\frac{-1}{2}(2k+1)} = \frac{1}{n!}\cdot\underbrace{\prod_{k=0}^{n-1}\left(\frac{-1}{2}(2k+1)\right)}_{=\left(\frac{-1}{2}\right)^n\prod_{k=0}^{n-1}(2k+1)} \\
&= \frac{1}{n!}\cdot\left(\frac{-1}{2}\right)^n\underbrace{\prod_{k=0}^{n-1}(2k+1)}_{\substack{=1\cdot 3\cdot 5\cdot\cdots\cdot(2(n-1)+1)\\ =1\cdot 3\cdot 5\cdot\cdots\cdot(2n-1)\\ (\text{since } 2(n-1)+1=2n-1)}} = \frac{1}{n!}\cdot\left(\frac{-1}{2}\right)^n\underbrace{(1\cdot 3\cdot 5\cdot\cdots\cdot(2n-1))}_{\substack{=(2n-1)\cdot(2n-3)\cdot\cdots\cdot 1\\ =\frac{(2n)!}{2^n n!}\\ (\text{by Exercise 3.2 \textbf{(a)}})}} \\
&= \frac{1}{n!}\cdot\left(\frac{-1}{2}\right)^n\cdot\frac{(2n)!}{2^n n!} = \left(\frac{-1}{2}\right)^n\cdot\underbrace{\frac{1}{2^n}}_{=\left(\frac{1}{2}\right)^n}\cdot\frac{(2n)!}{n!^2} = \left(\frac{-1}{2}\right)^n\cdot\left(\frac{1}{2}\right)^n\cdot\frac{(2n)!}{n!^2}.
\end{aligned}$$

Comparing this with

$$\left(\underbrace{\frac{-1}{4}}_{=\frac{-1}{2}\cdot\frac{1}{2}}\right)^n \underbrace{\binom{2n}{n}}_{\substack{=\frac{(2n)!}{n!^2} \\ \text{(by Lemma 7.7)}}} = \underbrace{\left(\frac{-1}{2}\cdot\frac{1}{2}\right)^n}_{=\left(\frac{-1}{2}\right)^n\cdot\left(\frac{1}{2}\right)^n} \cdot\frac{(2n)!}{n!^2} = \left(\frac{-1}{2}\right)^n \cdot \left(\frac{1}{2}\right)^n \cdot \frac{(2n)!}{n!^2},$$

we obtain $\binom{-1/2}{n} = \left(\frac{-1}{4}\right)^n \binom{2n}{n}$. This solves Exercise 3.2 **(b)**.

Second solution to Exercise 3.2 ***(b)****:* Let us solve Exercise 3.2 **(b)** by induction on n:

Induction base: Applying (227) to $m = 2 \cdot 0$, we obtain $\binom{2\cdot 0}{0} = 1$. But Proposition 3.3 **(a)** (applied to $-1/2$ instead of m) yields $\binom{-1/2}{0} = 1$. Comparing this with $\underbrace{\left(\frac{-1}{4}\right)^0}_{=1} \underbrace{\binom{2\cdot 0}{0}}_{=1} = 1$, we obtain $\binom{-1/2}{0} = \left(\frac{-1}{4}\right)^0 \binom{2\cdot 0}{0}$. In other words, Exercise 3.2 **(b)** holds for $n = 0$. This completes the induction base.

Induction step: Let m be a positive integer. Assume that Exercise 3.2 **(b)** holds for $n = m-1$. We must now prove that Exercise 3.2 **(b)** holds for $n = m$.

We have assumed that Exercise 3.2 **(b)** holds for $n = m-1$. In other words, we have

$$\binom{-1/2}{m-1} = \left(\frac{-1}{4}\right)^{m-1} \binom{2(m-1)}{m-1}. \tag{529}$$

We have $m - 1 \in \mathbb{N}$ (since m is a positive integer). Thus, Lemma 7.7 (applied to $n = m-1$) yields

$$\binom{2(m-1)}{m-1} = \frac{(2(m-1))!}{(m-1)!^2}. \tag{530}$$

Applying (37) to $n = m$, we obtain $m! = m \cdot (m-1)!$. Also, $m \geq 1$ (since m is a positive integer), so that $2m \geq 2$ and thus $2m - 1 \geq 2 - 1 = 1$. Hence, $2m - 1$ is a positive integer. Thus, (37) (applied to $n = 2m - 1$) yields

$$(2m-1)! = (2m-1) \cdot \left(\underbrace{(2m-1)-1}_{=2(m-1)}\right)! = (2m-1) \cdot (2(m-1))!.$$

Moreover, $2m$ is a positive integer (since $2m \geq 2 > 0$). Thus, (37) (applied to $n = 2m$) yields

$$(2m)! = (2m) \cdot \underbrace{(2m-1)!}_{=(2m-1)\cdot(2(m-1))!} = (2m) \cdot (2m-1) \cdot (2(m-1))!.$$

But Lemma 7.7 (applied to $n = m$) yields

$$\begin{aligned}
\binom{2m}{m} &= \frac{(2m)!}{m!^2} = \underbrace{(2m)!}_{=(2m)\cdot(2m-1)\cdot(2(m-1))!} / \left(\underbrace{m!}_{=m\cdot(m-1)!}\right)^2 \\
&= (2m)\cdot(2m-1)\cdot(2(m-1))!/\left(m\cdot(m-1)!\right)^2 \\
&= \frac{(2m)\cdot(2m-1)\cdot(2(m-1))!}{\left(m\cdot(m-1)!\right)^2} = \underbrace{\frac{2m}{m}}_{=2}\cdot\frac{2m-1}{m}\cdot\underbrace{\frac{(2(m-1))!}{(m-1)!^2}}_{\substack{=\binom{2(m-1)}{m-1} \\ \text{(by (530))}}} \\
&= 2\cdot\frac{2m-1}{m}\cdot\binom{2(m-1)}{m-1}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\underbrace{\left(\frac{-1}{4}\right)^m}_{=\left(\frac{-1}{4}\right)^{m-1}\cdot\frac{-1}{4}} \underbrace{\binom{2m}{m}}_{=2\cdot\frac{2m-1}{m}\cdot\binom{2(m-1)}{m-1}} \\
&= \left(\frac{-1}{4}\right)^{m-1}\cdot\underbrace{\frac{-1}{4}\cdot 2}_{=\frac{-1}{2}}\cdot\frac{2m-1}{m}\cdot\binom{2(m-1)}{m-1} \\
&= \left(\frac{-1}{4}\right)^{m-1}\cdot\frac{-1}{2}\cdot\frac{2m-1}{m}\cdot\binom{2(m-1)}{m-1}.
\end{aligned}$$

Comparing this with

$$\begin{aligned}
\binom{-1/2}{m} &= \underbrace{\frac{-1/2-m+1}{m}}_{=\frac{-1}{2}\cdot\frac{2m-1}{m}} \underbrace{\binom{-1/2}{m-1}}_{\substack{=\left(\frac{-1}{4}\right)^{m-1}\binom{2(m-1)}{m-1} \\ \text{(by (529))}}} \qquad \text{(by Lemma 7.6 (applied to } q = -1/2\text{))} \\
&= \frac{-1}{2}\cdot\frac{2m-1}{m}\cdot\left(\frac{-1}{4}\right)^{m-1}\binom{2(m-1)}{m-1} \\
&= \left(\frac{-1}{4}\right)^{m-1}\cdot\frac{-1}{2}\cdot\frac{2m-1}{m}\cdot\binom{2(m-1)}{m-1},
\end{aligned}$$

we obtain $\dbinom{-1/2}{m} = \left(\dfrac{-1}{4}\right)^m \dbinom{2m}{m}$. In other words, Exercise 3.2 **(b)** holds for $n = m$. This completes the induction step. Thus, the induction proof of Exercise 3.2 **(b)** is complete.

(c) *First solution to Exercise 3.2 **(c)**:* We have

$$(3n)! = 1 \cdot 2 \cdot \cdots \cdot (3n) = \prod_{k \in \{1,2,\ldots,3n\}} k$$
$$= \left(\prod_{\substack{k \in \{1,2,\ldots,3n\}; \\ k \equiv 0 \operatorname{mod} 3}} k \right) \left(\prod_{\substack{k \in \{1,2,\ldots,3n\}; \\ k \equiv 1 \operatorname{mod} 3}} k \right) \left(\prod_{\substack{k \in \{1,2,\ldots,3n\}; \\ k \equiv 2 \operatorname{mod} 3}} k \right)$$

(here, we have split the product $\prod_{k \in \{1,2,\ldots,3n\}} k$ into three smaller products, because each $k \in \{1, 2, \ldots, 3n\}$ must satisfy exactly one of the three conditions $k \equiv 0 \operatorname{mod} 3$, $k \equiv 1 \operatorname{mod} 3$ and $k \equiv 2 \operatorname{mod} 3$). Thus,

$$(3n)! = \underbrace{\left(\prod_{\substack{k \in \{1,2,\ldots,3n\}; \\ k \equiv 0 \operatorname{mod} 3}} k \right)}_{\substack{= 3 \cdot 6 \cdot 9 \cdot \cdots \cdot (3n) \\ = \prod_{i=1}^{n} (3i) \\ = 3^n \prod_{i=1}^{n} i}} \underbrace{\left(\prod_{\substack{k \in \{1,2,\ldots,3n\}; \\ k \equiv 1 \operatorname{mod} 3}} k \right)}_{\substack{= 1 \cdot 4 \cdot 7 \cdot \cdots \cdot (3n-2) \\ = \prod_{i=0}^{n-1} (3i+1)}} \underbrace{\left(\prod_{\substack{k \in \{1,2,\ldots,3n\}; \\ k \equiv 2 \operatorname{mod} 3}} k \right)}_{\substack{= 2 \cdot 5 \cdot 8 \cdot \cdots \cdot (3n-1) \\ = \prod_{i=0}^{n-1} (3i+2)}}$$
$$= 3^n \underbrace{\left(\prod_{i=1}^{n} i \right)}_{= n!} \left(\prod_{i=0}^{n-1} (3i+1) \right) \left(\prod_{i=0}^{n-1} (3i+2) \right)$$
$$= 3^n n! \left(\prod_{i=0}^{n-1} (3i+1) \right) \left(\prod_{i=0}^{n-1} (3i+2) \right). \tag{531}$$

On the other hand, for each $g \in \mathbb{Z}$, we have

$$\begin{aligned}
\dbinom{-g/3}{n} &= \frac{(-g/3)(-g/3-1)\cdots(-g/3-n+1)}{n!} \\
&\qquad \text{(by (226) (applied to } m = -g/3\text{))} \\
&= \frac{1}{n!} \cdot \underbrace{(-g/3)(-g/3-1)\cdots(-g/3-n+1)}_{=\prod_{i=0}^{n-1}(-g/3-i)} \\
&= \frac{1}{n!} \cdot \prod_{i=0}^{n-1} \underbrace{(-g/3-i)}_{=\frac{-1}{3}(3i+g)} = \frac{1}{n!} \cdot \underbrace{\prod_{i=0}^{n-1} \left(\frac{-1}{3}(3i+g)\right)}_{=\left(\frac{-1}{3}\right)^n \prod_{i=0}^{n-1}(3i+g)} \\
&= \frac{1}{n!} \cdot \left(\frac{-1}{3}\right)^n \prod_{i=0}^{n-1} (3i+g) . \qquad (532)
\end{aligned}$$

Now,

$$\begin{aligned}
&\underbrace{(3^n n!)^3}_{=3^n n! \cdot 3^n n! \cdot 3^n n!} \underbrace{\dbinom{-1/3}{n}}_{\substack{=\frac{1}{n!}\cdot\left(\frac{-1}{3}\right)^n \prod_{i=0}^{n-1}(3i+1) \\ \text{(by (532), applied to } g=1)}} \underbrace{\dbinom{-2/3}{n}}_{\substack{=\frac{1}{n!}\cdot\left(\frac{-1}{3}\right)^n \prod_{i=0}^{n-1}(3i+2) \\ \text{(by (532), applied to } g=2)}} \\
&= 3^n n! \cdot 3^n n! \cdot 3^n n! \cdot \left(\frac{1}{n!} \cdot \left(\frac{-1}{3}\right)^n \prod_{i=0}^{n-1}(3i+1)\right) \cdot \left(\frac{1}{n!} \cdot \left(\frac{-1}{3}\right)^n \prod_{i=0}^{n-1}(3i+2)\right) \\
&= \underbrace{3^n \cdot 3^n \cdot 3^n \cdot \left(\frac{-1}{3}\right)^n \left(\frac{-1}{3}\right)^n}_{=\left(3\cdot 3\cdot 3\cdot\frac{-1}{3}\cdot\frac{-1}{3}\right)^n} n! \left(\prod_{i=0}^{n-1}(3i+1)\right)\left(\prod_{i=0}^{n-1}(3i+2)\right) \\
&= \left(\underbrace{3\cdot 3\cdot 3\cdot \frac{-1}{3}\cdot\frac{-1}{3}}_{=3}\right)^n n! \left(\prod_{i=0}^{n-1}(3i+1)\right)\left(\prod_{i=0}^{n-1}(3i+2)\right) \\
&= 3^n n! \left(\prod_{i=0}^{n-1}(3i+1)\right)\left(\prod_{i=0}^{n-1}(3i+2)\right).
\end{aligned}$$

Comparing this with (531), we obtain $(3^n n!)^3 \dbinom{-1/3}{n}\dbinom{-2/3}{n} = (3n)!$. Dividing this equality by $(3^n n!)^3$, we find $\dbinom{-1/3}{n}\dbinom{-2/3}{n} = \dfrac{(3n)!}{(3^n n!)^3}$. This solves Exercise 3.2 **(c)**.

Second solution to Exercise 3.2 ***(c)****:* Let us solve Exercise 3.2 **(c)** by induction on n:

Induction base: Proposition 3.3 **(a)** (applied to $m = -1/3$) yields $\dbinom{-1/3}{0} = 1$. Also, Proposition 3.3 **(a)** (applied to $m = -2/3$) yields $\dbinom{-2/3}{0} = 1$. On the other hand,

$$\frac{(3 \cdot 0)!}{\left(3^0 0!\right)^3} = \left(\underbrace{3 \cdot 0}_{=0}\right)! / \left(\underbrace{3^0}_{=1} \underbrace{0!}_{=1}\right)^3 = 0!/1^3 = 0! = 1.$$

Comparing this with

$$\underbrace{\dbinom{-1/3}{0}}_{=1} \underbrace{\dbinom{-2/3}{0}}_{=1} = 1,$$

we obtain $\dbinom{-1/3}{0} \dbinom{-2/3}{0} = \dfrac{(3 \cdot 0)!}{\left(3^0 0!\right)^3}$. In other words, Exercise 3.2 **(c)** holds for $n = 0$. This completes the induction base.

Induction step: Let m be a positive integer. Assume that Exercise 3.2 **(c)** holds for $n = m - 1$. We must now prove that Exercise 3.2 **(c)** holds for $n = m$.

We have assumed that Exercise 3.2 **(c)** holds for $n = m - 1$. In other words, we have

$$\dbinom{-1/3}{m-1} \dbinom{-2/3}{m-1} = \frac{(3(m-1))!}{\left(3^{m-1}(m-1)!\right)^3}. \tag{533}$$

We have $m - 1 \in \mathbb{N}$ (since m is a positive integer). Hence, $3(m-1) \in \mathbb{N}$. The definition of $(3(m-1))!$ thus yields

$$(3(m-1))! = 1 \cdot 2 \cdot \cdots \cdot (3(m-1)) = 1 \cdot 2 \cdot \cdots \cdot (3m-3) \tag{534}$$

(since $3(m-1) = 3m - 3$).

The definition of $(3m)!$ yields

$$\begin{aligned} (3m)! &= 1 \cdot 2 \cdot \cdots \cdot (3m) = \underbrace{(1 \cdot 2 \cdot \cdots \cdot (3m-3))}_{\substack{=(3(m-1))! \\ \text{(by (534))}}} \cdot ((3m-2) \cdot (3m-1) \cdot (3m)) \\ &= (3(m-1))! \cdot ((3m-2) \cdot (3m-1) \cdot (3m)). \end{aligned}$$

Applying (37) to $n = m$, we obtain $m! = m \cdot (m-1)!$.

Now,

$$\begin{aligned}
\frac{(3m)!}{(3^m m!)^3} &= \underbrace{(3m)!}_{=(3(m-1))!\cdot((3m-2)\cdot(3m-1)\cdot(3m))} / \left(\underbrace{3^m}_{=3\cdot 3^{m-1}} \underbrace{m!}_{=m\cdot(m-1)!}\right)^3 \\
&= (3(m-1))! \cdot ((3m-2)\cdot(3m-1)\cdot(3m)) / \underbrace{\left(3\cdot 3^{m-1} m \cdot (m-1)!\right)^3}_{=3^3 m^3 \left(3^{m-1}(m-1)!\right)^3} \\
&= (3(m-1))! \cdot ((3m-2)\cdot(3m-1)\cdot(3m)) / \left(3^3 m^3 \left(3^{m-1}(m-1)!\right)^3\right) \\
&= \frac{(3(m-1))! \cdot ((3m-2)\cdot(3m-1)\cdot(3m))}{3^3 m^3 \left(3^{m-1}(m-1)!\right)^3} \\
&= \frac{(3(m-1))!}{\left(3^{m-1}(m-1)!\right)^3} \cdot \underbrace{\frac{(3m-2)\cdot(3m-1)\cdot(3m)}{3^3 m^3}}_{=\frac{3m-2}{3m}\cdot\frac{3m-1}{3m}} \\
&= \frac{(3(m-1))!}{\left(3^{m-1}(m-1)!\right)^3} \cdot \frac{3m-2}{3m} \cdot \frac{3m-1}{3m}. \qquad (535)
\end{aligned}$$

On the other hand, Lemma 7.6 (applied to $q = -1/3$) yields

$$\binom{-1/3}{m} = \underbrace{\frac{-1/3-m+1}{m}}_{=-\frac{3m-2}{3m}} \binom{-1/3}{m-1} = -\frac{3m-2}{3m}\binom{-1/3}{m-1}. \qquad (536)$$

Also, Lemma 7.6 (applied to $q = -2/3$) yields

$$\binom{-2/3}{m} = \underbrace{\frac{-2/3-m+1}{m}}_{=-\frac{3m-1}{3m}} \binom{-2/3}{m-1} = -\frac{3m-1}{3m}\binom{-2/3}{m-1}. \qquad (537)$$

Multiplying the two equalities (536) and (537), we obtain

$$\begin{aligned}
\binom{-1/3}{m}\binom{-2/3}{m} &= \left(-\frac{3m-2}{3m}\binom{-1/3}{m-1}\right)\left(-\frac{3m-1}{3m}\binom{-2/3}{m-1}\right) \\
&= \underbrace{\binom{-1/3}{m-1}\binom{-2/3}{m-1}}_{\substack{=\frac{(3(m-1))!}{\left(3^{m-1}(m-1)!\right)^3} \\ \text{(by (533))}}} \cdot \frac{3m-2}{3m} \cdot \frac{3m-1}{3m} \\
&= \frac{(3(m-1))!}{\left(3^{m-1}(m-1)!\right)^3} \cdot \frac{3m-2}{3m} \cdot \frac{3m-1}{3m} = \frac{(3m)!}{(3^m m!)^3}
\end{aligned}$$

(by (535)). In other words, Exercise 3.2 **(c)** holds for $n = m$. This completes the induction step. Thus, the induction proof of Exercise 3.2 **(c)** is complete. □

7.12.2. A more general formula

Parts **(b)** and **(c)** of Exercise 3.2 can be generalized as follows:

Theorem 7.8. Let h be a positive integer. Let $n \in \mathbb{N}$. Then,

$$\prod_{g=1}^{h-1} \binom{-g/h}{n} = \left(\frac{-1}{h}\right)^{n(h-1)} \cdot \frac{(hn)!}{h^n n!^h}.$$

Exercise 3.2 **(b)** follows from Theorem 7.8 (applied to $h = 2$), after some simple transformations (using Lemma 7.7). Exercise 3.2 **(c)** follows from Theorem 7.8 (applied to $h = 3$).

We are going to prove Theorem 7.8 in a way that is similar to our second solutions of parts **(b)** and **(c)** of Exercise 3.2:

Proof of Theorem 7.8. We shall prove Theorem 7.8 by induction on n:

Induction base: Using $0(h-1) = 0$, $h \cdot 0 = 0$ and $0! = 1$, we obtain

$$\begin{aligned}\left(\frac{-1}{h}\right)^{0(h-1)} \cdot \frac{(h \cdot 0)!}{h^0 0!^h} &= \underbrace{\left(\frac{-1}{h}\right)^0}_{=1} \cdot \frac{0!}{h^0 \cdot 1^h} = \frac{0!}{h^0 \cdot 1^h} = \frac{1}{1 \cdot 1} \\ &\qquad \left(\text{since } 0! = 1 \text{ and } h^0 = 1 \text{ and } 1^h = 1\right) \\ &= 1.\end{aligned}$$

Comparing this with

$$\prod_{g=1}^{h-1} \underbrace{\binom{-g/h}{0}}_{\substack{=1 \\ \text{(by Proposition 3.3 **(a)** (applied to } m=-g/h))}} = \prod_{g=1}^{h-1} 1 = 1,$$

we obtain $\prod_{g=1}^{h-1} \binom{-g/h}{0} = \left(\frac{-1}{h}\right)^{0(h-1)} \cdot \frac{(h \cdot 0)!}{h^0 0!^h}$. In other words, Theorem 7.8 holds for $n = 0$. This completes the induction base.

Induction step: Let m be a positive integer. Assume that Theorem 7.8 holds for $n = m - 1$. We must now prove that Theorem 7.8 holds for $n = m$.

We have $h > 0$ (since h is a positive integer) and $m > 0$ (since m is a positive integer). Thus, $hm > 0$, so that $hm \neq 0$.

We have assumed that Theorem 7.8 holds for $n = m-1$. In other words, we have

$$\prod_{g=1}^{h-1} \binom{-g/h}{m-1} = \left(\frac{-1}{h}\right)^{(m-1)(h-1)} \cdot \frac{(h(m-1))!}{h^{m-1}(m-1)!^h}. \tag{538}$$

We have $m-1 \in \mathbb{N}$ (since m is a positive integer) and $h \in \mathbb{N}$ (since h is a positive integer). Thus, $h(m-1) \in \mathbb{N}$. The definition of $(h(m-1))!$ thus yields

$$(h(m-1))! = 1 \cdot 2 \cdot \cdots \cdot (h(m-1)) = \prod_{i=1}^{h(m-1)} i. \tag{539}$$

We have $h(m-1) \in \mathbb{N}$ and thus $0 \leq h(m-1)$. Also, $hm - h(m-1) = h \geq 0$ (since $h \in \mathbb{N}$), so that $h(m-1) \leq hm$. The definition of $(hm)!$ yields

$$\begin{aligned}(hm)! &= 1 \cdot 2 \cdot \cdots \cdot (hm) = \prod_{i=1}^{hm} i = \underbrace{\left(\prod_{i=1}^{h(m-1)} i\right)}_{\substack{=(h(m-1))! \\ \text{(by (539))}}} \left(\prod_{i=h(m-1)+1}^{hm} i\right) \\ &\qquad (\text{since } 0 \leq h(m-1) \leq hm) \\ &= (h(m-1))! \cdot \left(\prod_{i=h(m-1)+1}^{hm} i\right).\end{aligned} \tag{540}$$

We have $h(m-1)+1 = (hm-h)+1$ and $hm = (hm-h)+h$. Thus,

$$\prod_{i=h(m-1)+1}^{hm} i = \prod_{i=(hm-h)+1}^{(hm-h)+h} i = \prod_{g=1}^{h} (g + (hm-h))$$

(here, we have substituted $g + (hm-h)$ for i in the product). Thus,

$$\begin{aligned}\prod_{i=h(m-1)+1}^{hm} i &= \prod_{g=1}^{h} (g + (hm-h)) = \left(\prod_{g=1}^{h-1} (g + (hm-h))\right) \cdot \underbrace{(h + (hm-h))}_{=hm} \\ &\qquad \left(\begin{array}{c}\text{here, we have split off the factor for } g = h \text{ from} \\ \text{the product (since } h > 0)\end{array}\right) \\ &= \left(\prod_{g=1}^{h-1} (g + (hm-h))\right) \cdot hm.\end{aligned}$$

We can divide this equality by hm (because $hm \neq 0$), and thus obtain

$$\frac{1}{hm} \prod_{i=h(m-1)+1}^{hm} i = \prod_{g=1}^{h-1} (g + (hm-h)). \tag{541}$$

Now,

$$\begin{aligned}
&\prod_{g=1}^{h-1} \underbrace{\frac{-g/h-m+1}{m}}_{=\left(-\frac{1}{h}\right)\frac{1}{m}(g+(hm-h))} \\
&= \prod_{g=1}^{h-1} \left(\left(-\frac{1}{h}\right) \frac{1}{m} \left(g+(hm-h)\right) \right) \\
&= \underbrace{\left(\prod_{g=1}^{h-1} \left(-\frac{1}{h}\right) \right)}_{=\left(-\frac{1}{h}\right)^{h-1}} \underbrace{\left(\prod_{g=1}^{h-1} \frac{1}{m} \right)}_{=\left(\frac{1}{m}\right)^{h-1} = \frac{1}{m^{h-1}}} \underbrace{\left(\prod_{g=1}^{h-1} \left(g+(hm-h)\right) \right)}_{\substack{=\frac{1}{hm} \prod_{i=h(m-1)+1}^{hm} i \\ \text{(by (541))}}} \\
&= \left(-\frac{1}{h}\right)^{h-1} \frac{1}{m^{h-1}} \cdot \frac{1}{hm} \prod_{i=h(m-1)+1}^{hm} i = \left(-\frac{1}{h}\right)^{h-1} \underbrace{\frac{1}{m^{h-1}m}}_{\substack{=\frac{1}{m^h} \\ \text{(since } m^{h-1}m=m^h\text{)}}} \cdot \frac{1}{h} \prod_{i=h(m-1)+1}^{hm} i \\
&= \left(-\frac{1}{h}\right)^{h-1} \frac{1}{m^h} \cdot \frac{1}{h} \prod_{i=h(m-1)+1}^{hm} i. \qquad (542)
\end{aligned}$$

We have

$$\begin{aligned}
&\prod_{g=1}^{h-1} \underbrace{\binom{-g/h}{m}}_{\substack{=\frac{-g/h-m+1}{m}\binom{-g/h}{m-1} \\ \text{(by Lemma 7.6 (applied to } q=-g/h\text{))}}} \\
&= \prod_{g=1}^{h-1} \left(\frac{-g/h-m+1}{m} \binom{-g/h}{m-1} \right) \\
&= \underbrace{\left(\prod_{g=1}^{h-1} \frac{-g/h-m+1}{m} \right)}_{\substack{=\left(-\frac{1}{h}\right)^{h-1} \frac{1}{m^h} \cdot \frac{1}{h} \prod_{i=h(m-1)+1}^{hm} i \\ \text{(by (542))}}} \underbrace{\left(\prod_{g=1}^{h-1} \binom{-g/h}{m-1} \right)}_{\substack{=\left(\frac{-1}{h}\right)^{(m-1)(h-1)} \cdot \frac{(h(m-1))!}{h^{m-1}(m-1)!^h} \\ \text{(by (538))}}} \\
&= \left(-\frac{1}{h}\right)^{h-1} \frac{1}{m^h} \cdot \frac{1}{h} \left(\prod_{i=h(m-1)+1}^{hm} i \right) \cdot \left(\frac{-1}{h}\right)^{(m-1)(h-1)} \cdot \frac{(h(m-1))!}{h^{m-1}(m-1)!^h} \\
&= \underbrace{\left(-\frac{1}{h}\right)^{h-1} \cdot \left(\frac{-1}{h}\right)^{(m-1)(h-1)}}_{\substack{=\left(-\frac{1}{h}\right)^{(h-1)+(m-1)(h-1)} \\ =\left(-\frac{1}{h}\right)^{m(h-1)} \\ \text{(since } (h-1)+(m-1)(h-1)=m(h-1)\text{)}}} \cdot \left(\prod_{i=h(m-1)+1}^{hm} i \right) \cdot \frac{1}{m^h} \cdot \frac{1}{h} \frac{(h(m-1))!}{h^{m-1}(m-1)!^h} \\
&= \left(-\frac{1}{h}\right)^{m(h-1)} \cdot \left(\prod_{i=h(m-1)+1}^{hm} i \right) \cdot \frac{1}{m^h} \cdot \frac{1}{h} \frac{(h(m-1))!}{h^{m-1}(m-1)!^h} \\
&= \left(-\frac{1}{h}\right)^{m(h-1)} \cdot \underbrace{(h(m-1))! \cdot \left(\prod_{i=h(m-1)+1}^{hm} i \right)}_{\substack{=(hm)! \\ \text{(by (540))}}} \cdot \frac{1}{m^h} \cdot \underbrace{\frac{1}{hh^{m-1}(m-1)!^h}}_{\substack{=\frac{1}{h^m(m-1)!^h} \\ \text{(since } hh^{m-1}=h^m\text{)}}} \\
&= \left(-\frac{1}{h}\right)^{m(h-1)} \cdot (hm)! \cdot \frac{1}{m^h} \cdot \frac{1}{h^m(m-1)!^h} \\
&= \left(-\frac{1}{h}\right)^{m(h-1)} \cdot \frac{(hm)!}{h^m(m-1)!^h m^h}. \qquad (543)
\end{aligned}$$

Applying (37) to $n = m$, we obtain $m! = m \cdot (m-1)!$. Thus,

$$m!^h = (m \cdot (m-1)!)^h = m^h \cdot (m-1)!^h.$$

Hence,

$$\left(-\frac{1}{h}\right)^{m(h-1)} \cdot \frac{(hm)!}{h^m m!^h} = \left(-\frac{1}{h}\right)^{m(h-1)} \cdot \frac{(hm)!}{h^m m^h \cdot (m-1)!^h} = \left(-\frac{1}{h}\right)^{m(h-1)} \cdot \frac{(hm)!}{h^m (m-1)!^h m^h}.$$

Comparing this with (543), we obtain

$$\prod_{g=1}^{h-1} \binom{-g/h}{m} = \left(-\frac{1}{h}\right)^{m(h-1)} \cdot \frac{(hm)!}{h^m m!^h}.$$

In other words, Theorem 7.8 holds for $n = m$. This completes the induction step. Thus, the induction proof of Theorem 7.8 is complete. □

7.13. Solution to Exercise 3.3

Let us start by proving part **(a)** of Exercise 3.3:

Lemma 7.9. Let $n \in \mathbb{N}$ and $q \in \mathbb{Q}$. Then,

$$\sum_{r=0}^{n} \binom{r+q}{r} = \binom{n+q+1}{n}.$$

Proof of Lemma 7.9. Let us forget that n is fixed. Now, let us prove Lemma 7.9 by induction on n:

Induction base: Proposition 3.3 **(a)** (applied to $m = 0 + q$) yields $\binom{0+q}{0} = 1$. Similarly, $\binom{0+q+1}{0} = 1$. Comparing this with $\sum_{r=0}^{0} \binom{r+q}{r} = \binom{0+q}{0} = 1$, we obtain $\sum_{r=0}^{0} \binom{r+q}{r} = \binom{0+q+1}{0}$. In other words, Lemma 7.9 holds for $n = 0$. This completes the induction base.

Induction step: Let m be a positive integer. Assume that Lemma 7.9 holds for $n = m - 1$. We must prove that Lemma 7.9 holds for $n = m$.

We have assumed that Lemma 7.9 holds for $n = m - 1$. In other words, we have

$$\sum_{r=0}^{m-1} \binom{r+q}{r} = \binom{(m-1)+q+1}{m-1}.$$

We have $m \in \{1, 2, 3, \ldots\}$ (since m is a positive integer). Thus, Proposition 3.11 (applied to $m + q + 1$ and m instead of m and n) yields

$$\dbinom{m+q+1}{m} = \dbinom{(m+q+1)-1}{m} + \dbinom{(m+q+1)-1}{m-1} = \dbinom{m+q}{m} + \dbinom{m+q}{m-1} \qquad (544)$$

(since $(m + q + 1) - 1 = m + q$).

On the other hand, $m \in \{0, 1, \ldots, m\}$ (since m is a positive integer). Thus, we can split off the addend for $r = m$ from the sum $\sum\limits_{r=0}^{m} \dbinom{r+q}{r}$. We thus obtain

$$\sum_{r=0}^{m} \dbinom{r+q}{r} = \underbrace{\sum_{r=0}^{m-1} \dbinom{r+q}{r}}_{\substack{=\dbinom{(m-1)+q+1}{m-1} \\ =\dbinom{m+q}{m-1} \\ \text{(since } (m-1)+q+1=m+q)}} + \dbinom{m+q}{m}$$

$$= \dbinom{m+q}{m-1} + \dbinom{m+q}{m} = \dbinom{m+q}{m} + \dbinom{m+q}{m-1} = \dbinom{m+q+1}{m}$$

(by (544)). In other words, Lemma 7.9 holds for $n = m$. This completes the induction step. Thus, Lemma 7.9 is proven by induction. □

Next, let us show the two equalities in Exercise 3.3 **(b)** separately:

Lemma 7.10. Let $n \in \{-1, 0, 1, \ldots\}$ and $k \in \mathbb{N}$. Then:

(a) We have

$$\sum_{i=k}^{n} \dbinom{i}{k} = \dbinom{n+1}{k+1}.$$

(b) We have

$$\sum_{i=0}^{n} \dbinom{i}{k} = \sum_{i=k}^{n} \dbinom{i}{k}.$$

Proof of Lemma 7.10. We have $n \in \{-1, 0, 1, \ldots\}$, so that $n \geq -1$ and therefore $n + 1 \geq 0$. Hence, $n + 1 \in \mathbb{N}$.

(a) If $n < k$, then Lemma 7.10 **(a)** holds[300]. Hence, for the rest of this proof of Lemma 7.10 **(a)**, we can WLOG assume that we don't have $n < k$. Assume this.

[300]*Proof.* Assume that $n < k$. We must show that Lemma 7.10 **(a)** holds.

Recall that $n + 1 \in \mathbb{N}$. Also, $k + 1 > k \geq 0$ (since $k \in \mathbb{N}$) and thus $k + 1 \in \mathbb{N}$. Finally, We have $n \geq k$ (since we don't have $n < k$). Hence, $n \geq k \geq 0$ (since $k \in \mathbb{N}$), so that $n \in \mathbb{N}$. Also, from $n \geq k$, we obtain $n - k \geq 0$, so that $n - k \in \mathbb{N}$.

Lemma 7.9 (applied to $n - k$ and k instead of n and q) yields

$$\sum_{r=0}^{n-k} \binom{r+k}{r} = \binom{(n-k)+k+1}{n-k} = \binom{n+1}{n-k} \tag{545}$$

(since $(n-k)+k+1 = n+1$).

We have $n+1 \in \mathbb{N}$ and $k+1 \in \mathbb{N}$ (since $k \in \mathbb{N}$). Also, $\underbrace{n}_{\geq k} + 1 \geq k+1$. Thus, Proposition 3.8 (applied to $n+1$ and $k+1$ instead of n and k) yields

$$\binom{n+1}{k+1} = \binom{n+1}{(n+1)-(k+1)} = \binom{n+1}{n-k}$$

(since $(n+1)-(k+1) = n-k$). Comparing this equality with (545), we obtain

$$\binom{n+1}{k+1} = \sum_{r=0}^{n-k} \binom{r+k}{r} = \sum_{i=k}^{n} \binom{(i-k)+k}{i-k} \tag{546}$$

(here, we have substituted $i-k$ for r in the sum).

But each $i \in \{k, k+1, \ldots, n\}$ satisfies

$$\binom{(i-k)+k}{i-k} = \binom{i}{k}. \tag{547}$$

[*Proof of (547):* Let $i \in \{k, k+1, \ldots, n\}$. Thus, $i \geq k$ and $i \in \{k, k+1, \ldots, n\} \subseteq \mathbb{N}$. Hence, Proposition 3.8 (applied to i and k instead of m and n) yields $\binom{i}{k} = \binom{i}{i-k}$. Now, $(i-k)+k = i$, so that $\binom{(i-k)+k}{i-k} = \binom{i}{i-k} = \binom{i}{k}$. This proves (547).]

$\underbrace{n}_{<k} + 1 < k+1$. Hence, Proposition 3.6 (applied to $n+1$ and $k+1$ instead of m and n) yields $\binom{n+1}{k+1} = 0$. Comparing this with

$$\sum_{i=k}^{n} \binom{i}{k} = (\text{empty sum}) \qquad (\text{since } n < k)$$
$$= 0,$$

we obtain $\sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$. In other words, Lemma 7.10 **(a)** holds. Thus, we have shown that if $n < k$, then Lemma 7.10 **(a)** holds.

Now, (546) becomes

$$\binom{n+1}{k+1} = \sum_{i=k}^{n} \underbrace{\binom{(i-k)+k}{i-k}}_{\substack{=\binom{i}{k} \\ \text{(by (547))}}} = \sum_{i=k}^{n} \binom{i}{k}.$$

In other words, $\sum\limits_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$. This proves Lemma 7.10 **(a)**.

(b) If $n < k$, then Lemma 7.10 **(b)** holds[301]. Hence, for the rest of this proof, we can WLOG assume that we don't have $n < k$. Assume this.

Let $i \in \{0, 1, \ldots, k-1\}$. Thus, $i \leq k - 1 < k$ and $i \in \{0, 1, \ldots, k-1\} \subseteq \mathbb{N}$. Hence, Proposition 3.6 (applied to i and k instead of m and n) yields $\binom{i}{k} = 0$.

Now, forget that we fixed i. We thus have proven the equality $\binom{i}{k} = 0$ for each $i \in \{0, 1, \ldots, k-1\}$. Adding up these equalities for all $i \in \{0, 1, \ldots, k-1\}$, we obtain $\sum\limits_{i=0}^{k-1} \binom{i}{k} = \sum\limits_{i=0}^{k-1} 0 = 0$.

From $k \in \mathbb{N}$, we obtain $k \geq 0$, so that $0 \leq k \leq n$ (since $n \geq k$ (because we don't have $n < k$)). Hence, we can split the sum $\sum\limits_{i=0}^{n} \binom{i}{k}$ at $i = k$. We thus obtain

$$\sum_{i=0}^{n} \binom{i}{k} = \underbrace{\sum_{i=0}^{k-1} \binom{i}{k}}_{=0} + \sum_{i=k}^{n} \binom{i}{k} = \sum_{i=k}^{n} \binom{i}{k}.$$

This proves Lemma 7.10 **(b)**. □

[301] *Proof.* Assume that $n < k$. We must show that Lemma 7.10 **(b)** holds.

Let $i \in \{0, 1, \ldots, n\}$. Thus, $i \leq n < k$ and $i \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$. Hence, Proposition 3.6 (applied to i and k instead of m and n) yields $\binom{i}{k} = 0$.

Now, forget that we fixed i. We thus have proven the equality $\binom{i}{k} = 0$ for each $i \in \{0, 1, \ldots, n\}$. Adding up these equalities for all $i \in \{0, 1, \ldots, n\}$, we obtain $\sum\limits_{i=0}^{n} \binom{i}{k} = \sum\limits_{i=0}^{n} 0 = 0$. Comparing this with

$$\sum_{i=k}^{n} \binom{i}{k} = (\text{empty sum}) \qquad (\text{since } n < k)$$
$$= 0,$$

we obtain $\sum\limits_{i=0}^{n} \binom{i}{k} = \sum\limits_{i=k}^{n} \binom{i}{k}$. In other words, Lemma 7.10 **(b)** holds. Thus, we have shown that if $n < k$, then Lemma 7.10 **(b)** holds.

Solution to Exercise 3.3. **(a)** Let $n \in \mathbb{N}$ and $q \in \mathbb{Q}$. Then, Lemma 7.9 yields

$$\sum_{r=0}^{n} \binom{r+q}{r} = \binom{n+q+1}{n}.$$

This solves Exercise 3.3 **(a)**.

(b) Let $n \in \{-1, 0, 1, \ldots\}$ and $k \in \mathbb{N}$. Lemma 7.10 **(b)** yields

$$\sum_{i=0}^{n} \binom{i}{k} = \sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$$

(by Lemma 7.10 **(a)**). This solves Exercise 3.3 **(b)**. □

7.14. Solution to Exercise 3.4

Exercise 3.4 asks us to prove Proposition 3.12.

First, let us handle two trivial cases of Proposition 3.12:

Lemma 7.11. Proposition 3.12 holds in the case when $n = 0$.

Proof of Lemma 7.11. Let $m \in \mathbb{N}$, and let S be an m-element set. Then, the set S has exactly one 0-element subset (namely, the empty set $\varnothing$). Thus, the number of all 0-element subsets of S is 1.

On the other hand, $\binom{m}{0} = 1$ (by (227)). Thus, $\binom{m}{0}$ is the number of all 0-element subsets of S (since the number of all 0-element subsets of S is 1).

Now, forget that we fixed m and S. We thus have shown that if $m \in \mathbb{N}$ and if S is an m-element set, then $\binom{m}{0}$ is the number of all 0-element subsets of S. In other words, Proposition 3.12 holds in the case when $n = 0$. This proves Lemma 7.11. □

Lemma 7.12. Proposition 3.12 holds in the case when $m = 0$.

Proof of Lemma 7.12. Let m, n and S be as in Proposition 3.12. Assume that $m = 0$. We then must prove that Proposition 3.12 holds for these m, n and S.

If $n = 0$, then this follows from Lemma 7.11. Hence, for the rest of this proof, we can WLOG assume that $n \neq 0$. Assume this.

From $n \neq 0$, we conclude that n is a positive integer. Hence, $n > 0 = m$, so that $m < n$. Thus, (231) yields $\binom{m}{n} = 0$.

But S is an m-element set, i.e., a 0-element set (since $m = 0$). In other words, $S = \varnothing$. Hence, S has no n-element subsets[302]. In other words, the number of all n-element subsets of S is 0. Since $\binom{m}{n}$ is also 0, this shows that $\binom{m}{n}$ is the number

[302]*Proof.* Assume the contrary. Thus, S has an n-element subset. Let Q be such a subset. Thus, $Q \subseteq S$ and $|Q| = n$. Since $Q \subseteq S = \varnothing$, we have $Q = \varnothing$ and thus $|Q| = 0$. This contradicts $|Q| = n \neq 0$. This contradiction shows that our assumption was wrong, qed.

of all n-element subsets of S. Hence, Proposition 3.12 holds for our m, n and S. This proves Lemma 7.12. $\square$

Proof of Proposition 3.12. We shall prove Proposition 3.12 by induction over m:

Induction base: Lemma 7.12 shows that Proposition 3.12 holds in the case when $m = 0$. This completes the induction base.

Induction step: Let M be a positive integer. Assume that Proposition 3.12 holds for $m = M - 1$. We now must prove that Proposition 3.12 holds for $m = M$.

We have assumed that Proposition 3.12 holds for $m = M - 1$. In other words, if $n \in \mathbb{N}$, and if S is an $(M-1)$-element set, then

$$\binom{M-1}{n} \text{ is the number of all } n\text{-element subsets of } S. \tag{548}$$

Now, let $n \in \mathbb{N}$, and let S be an M-element set. We shall show that

$$\binom{M}{n} \text{ is the number of all } n\text{-element subsets of } S. \tag{549}$$

[*Proof of (549):* If $n = 0$, then (549) follows from Lemma 7.11[303]. Thus, for the rest of this proof of (549), we can WLOG assume that $n \neq 0$. Assume this.

Now, n is a positive integer (since $n \in \mathbb{N}$ and $n \neq 0$); thus, $n - 1 \in \mathbb{N}$.

Now, S is an M-element set; thus, $|S| = M > 0$. Hence, the set S is nonempty. In other words, there exists an $s \in S$. Fix such an s. Then, $|S \setminus \{s\}| = \underbrace{|S|}_{=M} - 1 = M - 1$. In other words, $S \setminus \{s\}$ is an $(M-1)$-element set. Hence, (548) (applied to $S \setminus \{s\}$ instead of S) yields that

$$\binom{M-1}{n} \text{ is the number of all } n\text{-element subsets of } S \setminus \{s\}. \tag{550}$$

Also, (548) (applied to $n-1$ and $S \setminus \{s\}$ instead of n and S) yields that

$$\binom{M-1}{n-1} \text{ is the number of all } (n-1)\text{-element subsets of } S \setminus \{s\}. \tag{551}$$

Now, the n-element subsets of S can be classified into two types: the ones that contain s, and the ones that don't. We can count them separately:

[303] *Proof.* Lemma 7.11 yields that Proposition 3.12 holds for $n = 0$. Hence, we can apply Proposition 3.12 to $n = 0$ and $m = M$. We thus obtain that $\binom{M}{0}$ is the number of all 0-element subsets of S. If $n = 0$, then this rewrites as follows: $\binom{M}{n}$ is the number of all n-element subsets of S. Hence, if $n = 0$, then (549) holds.

- The n-element subsets of S that contain s are in bijection with the $(n-1)$-element subsets of $S \setminus \{s\}$. More precisely: To each $(n-1)$-element subset U of $S \setminus \{s\}$, we can assign a unique n-element subset of S that contains s (namely, $U \cup \{s\}$); and this assignment is bijective (i.e., each n-element subset of S that contains s gets assigned to exactly one $(n-1)$-element subsets of $S \setminus \{s\}$). [304] Hence,

$$\begin{aligned}
&(\text{the number of all } n\text{-element subsets of } S \text{ that contain } s) \\
&= \left(\text{the number of all } (n-1)\text{-element subsets of } S \setminus \{s\}\right) \\
&= \binom{M-1}{n-1} \qquad (\text{by (551)}) .
\end{aligned} \tag{552}$$

- The n-element subsets of S that don't contain s are precisely the n-element subsets of $S \setminus \{s\}$ (because the subsets of S that don't contain s are precisely the subsets of $S \setminus \{s\}$). Hence,

$$\begin{aligned}
&(\text{the number of all } n\text{-element subsets of } S \text{ that don't contain } s) \\
&= (\text{the number of all } n\text{-element subsets of } S \setminus \{s\}) \\
&= \binom{M-1}{n} \qquad (\text{by (550)}) .
\end{aligned} \tag{553}$$

[304] Let me restate this in even more formal terms:

Let $\mathbf{A}$ be the set of all $(n-1)$-element subsets of $S \setminus \{s\}$. Let $\mathbf{B}$ be the set of all n-element subsets of S that contain s. Then, the map

$$\mathbf{A} \to \mathbf{B}, \qquad U \mapsto U \cup \{s\}$$

is well-defined and bijective.

(If you want to prove this formally, you need to prove two statements:

1. The map $\mathbf{A} \to \mathbf{B}$, $U \mapsto U \cup \{s\}$ is well-defined (i.e., we have $U \cup \{s\} \in \mathbf{B}$ for each $U \in \mathbf{A}$).
2. This map is bijective.

Proving the first statement is straightforward. The best way to prove the second statement is to show that the map $\mathbf{A} \to \mathbf{B}$, $U \mapsto U \cup \{s\}$ has an inverse – namely, the map $\mathbf{B} \to \mathbf{A}$, $V \mapsto V \setminus \{s\}$. Of course, you would also have to show that this latter map is well-defined, too.)

Now, every n-element subset of S either contains s or does not. Hence,

$$\begin{aligned}
&(\text{the number of all } n\text{-element subsets of } S) \\
&= \underbrace{(\text{the number of all } n\text{-element subsets of } S \text{ that contain } s)}_{\substack{=\dbinom{M-1}{n-1} \\ \text{(by (552))}}} \\
&\qquad + \underbrace{(\text{the number of all } n\text{-element subsets of } S \text{ that don't contain } s)}_{\substack{=\dbinom{M-1}{n} \\ \text{(by (553))}}} \\
&= \dbinom{M-1}{n-1} + \dbinom{M-1}{n}.
\end{aligned}$$

Compared with

$$\dbinom{M}{n} = \dbinom{M-1}{n-1} + \dbinom{M-1}{n} \qquad \text{(by (234), applied to } m = M\text{)},$$

this yields $(\text{the number of all } n\text{-element subsets of } S) = \dbinom{M}{n}$. Hence, (549) is proven.]

Now, forget that we fixed n and S. We thus have shown that every $n \in \mathbb{N}$ and every M-element set S satisfy (549). In other words, Proposition 3.12 holds for $m = M$. This completes the induction step. Thus, Proposition 3.12 is proven by induction. □

7.15. Solution to Exercise 3.5

Exercise 3.5 **(a)** is precisely the statement of Corollary 3.17 (with n renamed as k). Let me still give two proofs for it.

Solution to Exercise 3.5. **(a)** *First solution to Exercise 3.5* ***(a)****:* Let $k \in \mathbb{N}$. Then, Proposition 3.16 (applied to $m = -1$ and $n = k$) yields

$$\begin{aligned}
\dbinom{-1}{k} &= (-1)^k \dbinom{k-(-1)-1}{k} \\
&= (-1)^k \underbrace{\dbinom{k}{k}}_{\substack{=1 \\ \text{(by Proposition 3.9} \\ \text{(applied to } m=k))}} \qquad (\text{since } k-(-1)-1 = k) \\
&= (-1)^k .
\end{aligned}$$

This solves Exercise 3.5 **(a)**.

*Second solution to Exercise 3.5 **(a)**:* Let $k \in \mathbb{N}$. Then, the definition of $k!$ yields $k! = 1 \cdot 2 \cdot \cdots \cdot k$. But (226) (applied to -1 and k instead of m and n) yields

$$\begin{aligned}
\binom{-1}{k} &= \frac{(-1)(-2)\cdots(-1-k+1)}{k!} \\
&= \frac{(-1)(-2)\cdots(-k)}{k!} \qquad (\text{since } -1-k+1=-k) \\
&= \frac{1}{k!} \cdot \underbrace{(-1)(-2)\cdots(-k)}_{=(-1)^k\cdot(1\cdot 2\cdot\cdots\cdot k)} = \frac{1}{k!} \cdot (-1)^k \cdot \underbrace{(1\cdot 2\cdot\cdots\cdot k)}_{=k!} \\
&= \frac{1}{k!} \cdot (-1)^k \cdot k! = (-1)^k .
\end{aligned}$$

This solves Exercise 3.5 **(a)**.

(b) *First solution to Exercise 3.5 **(b)**:* Let $k \in \mathbb{N}$. Then, $k+1 \in \mathbb{N}$ and $k+1 \geq k$. Hence, Proposition 3.8 (applied to $m = k+1$ and $n = k$) yields

$$\begin{aligned}
\binom{k+1}{k} &= \binom{k+1}{(k+1)-k} = \binom{k+1}{1} \qquad (\text{since } (k+1)-k=1) \\
&= k+1 \qquad (\text{by Proposition 3.3 } \textbf{(b)} \text{ (applied to } m=k+1)) .
\end{aligned}$$

But Proposition 3.16 (applied to $m = -2$ and $n = k$) yields

$$\begin{aligned}
\binom{-2}{k} &= (-1)^k \binom{k-(-2)-1}{k} \\
&= (-1)^k \underbrace{\binom{k+1}{k}}_{=k+1} \qquad (\text{since } k-(-2)-1=k+1) \\
&= (-1)^k (k+1) .
\end{aligned}$$

This solves Exercise 3.5 **(b)**.

*Second solution to Exercise 3.5 **(b)**:* Let $k \in \mathbb{N}$. Then, the definition of $(k+1)!$ yields

$$(k+1)! = 1 \cdot 2 \cdot \cdots \cdot (k+1) = 1 \cdot (2 \cdot 3 \cdot \cdots \cdot (k+1)) = 2 \cdot 3 \cdot \cdots \cdot (k+1) .$$

On the other hand, applying (37) to $n = k+1$, we find

$$(k+1)! = (k+1) \cdot \left(\underbrace{(k+1)-1}_{=k} \right)! = (k+1) \cdot k!.$$

But (226) (applied to -2 and k instead of m and n) yields

$$\begin{aligned}\binom{-2}{k} &= \frac{(-2)(-3)\cdots(-2-k+1)}{k!} \\ &= \frac{(-2)(-3)\cdots(-(k+1))}{k!} \qquad (\text{since } -2-k+1 = -(k+1)) \\ &= \frac{1}{k!}\cdot\underbrace{(-2)(-3)\cdots(-(k+1))}_{=(-1)^k\cdot(2\cdot3\cdot\cdots\cdot(k+1))} = \frac{1}{k!}\cdot(-1)^k\cdot\underbrace{(2\cdot3\cdot\cdots\cdot(k+1))}_{=(k+1)!} \\ &= \frac{1}{k!}\cdot(-1)^k\cdot\underbrace{(k+1)!}_{=(k+1)\cdot k!} = \frac{1}{k!}\cdot(-1)^k\cdot(k+1)\cdot k! = (-1)^k(k+1).\end{aligned}$$

This solves Exercise 3.5 **(b)**.

(c) We shall give two solutions to Exercise 3.5 **(c)**: one by astutely transforming the left-hand side, and another by straightforward induction.

*First solution to Exercise 3.5 **(c)**:* Let $n \in \mathbb{N}$. Then, we can group the factors of the product $1!\cdot 2!\cdot\cdots\cdot(2n)!$ into pairs of successive factors. We thus obtain

$$\begin{aligned}&1!\cdot2!\cdot\cdots\cdot(2n)! \\ &= (1!\cdot2!)\cdot(3!\cdot4!)\cdot\cdots\cdot((2n-1)!\cdot(2n)!) = \prod_{i=1}^{n}\left((2i-1)!\cdot\underbrace{(2i)!}_{\substack{=2i\cdot(2i-1)! \\ \text{(by (37), applied} \\ \text{to } n=2i)}}\right) \\ &= \prod_{i=1}^{n}\underbrace{((2i-1)!\cdot 2i\cdot(2i-1)!)}_{=2i\cdot(2i-1)!^2} = \prod_{i=1}^{n}\left(2i\cdot(2i-1)!^2\right) \\ &= \underbrace{\left(\prod_{i=1}^{n}(2i)\right)}_{=2^n\prod_{i=1}^{n}i}\cdot\underbrace{\left(\prod_{i=1}^{n}\left((2i-1)!^2\right)\right)}_{=\left(\prod_{i=1}^{n}((2i-1)!)\right)^2} = 2^n\underbrace{\left(\prod_{i=1}^{n}i\right)}_{=1\cdot2\cdot\cdots\cdot n=n!}\cdot\left(\prod_{i=1}^{n}((2i-1)!)\right)^2 \\ &= 2^n n!\cdot\left(\prod_{i=1}^{n}((2i-1)!)\right)^2.\end{aligned}$$

Dividing both sides of this equality by $n!$, we find

$$\frac{1!\cdot2!\cdot\cdots\cdot(2n)!}{n!} = 2^n\cdot\left(\prod_{i=1}^{n}((2i-1)!)\right)^2.$$

This solves Exercise 3.5 **(c)**.

*Second solution to Exercise 3.5 **(c)**:* We shall solve Exercise 3.5 **(c)** by induction on n:

Induction base: Comparing

$$\frac{1! \cdot 2! \cdot \cdots \cdot (2 \cdot 0)!}{0!} = \underbrace{\frac{1}{0!}}_{=1} \cdot \underbrace{(1! \cdot 2! \cdot \cdots \cdot (2 \cdot 0)!)}_{\substack{=1! \cdot 2! \cdot \cdots \cdot 0! \\ =(\text{empty product})=1}} = 1$$

with

$$\underbrace{2^0}_{=1} \cdot \left(\underbrace{\prod_{i=1}^{0} ((2i-1)!)}_{=(\text{empty product})=1} \right)^2 = 1 \cdot 1^2 = 1,$$

we obtain $\frac{1! \cdot 2! \cdot \cdots \cdot (2 \cdot 0)!}{0!} = 2^0 \cdot \left(\prod_{i=1}^{0} ((2i-1)!) \right)^2$. In other words, Exercise 3.5 **(c)** holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Exercise 3.5 **(c)** holds for $n = m$. We must prove that Exercise 3.5 **(c)** holds for $n = m + 1$.

Clearly, $m + 1$ is a positive integer (since $m + 1 \geq 1 > 0$). Thus, applying (37) to $n = m + 1$, we find

$$(m+1)! = (m+1) \cdot \left(\underbrace{(m+1) - 1}_{=m} \right)! = (m+1) \cdot m!.$$

On the other hand, $2m + 2 \geq 2 > 0$, so that $2m + 2$ is a positive integer. Hence, applying (37) to $n = 2m + 2$, we find

$$(2m+2)! = (2m+2) \cdot \left(\underbrace{(2m+2) - 1}_{=2m+1} \right)! = (2m+2) \cdot (2m+1)!.$$

We have assumed that Exercise 3.5 **(c)** holds for $n = m$. In other words, we have

$$\frac{1! \cdot 2! \cdot \cdots \cdot (2m)!}{m!} = 2^m \cdot \left(\prod_{i=1}^{m} ((2i-1)!) \right)^2.$$

Now,

$$\begin{aligned}
&\frac{1!\cdot 2!\cdot \cdots \cdot (2(m+1))!}{(m+1)!}\\
&=\frac{1!\cdot 2!\cdot \cdots \cdot (2m+2)!}{(m+1)\cdot m!}\\
&\qquad \left(\text{since } 2(m+1)=2m+2 \text{ and } (m+1)!=(m+1)\cdot m!\right)\\
&=\frac{1}{(m+1)\cdot m!}\cdot \underbrace{(1!\cdot 2!\cdot \cdots \cdot (2m+2)!)}_{=(1!\cdot 2!\cdot\cdots\cdot(2m)!)\cdot(2m+1)!\cdot(2m+2)!}\\
&=\frac{1}{(m+1)\cdot m!}\cdot (1!\cdot 2!\cdot \cdots \cdot (2m)!)\cdot (2m+1)!\cdot \underbrace{(2m+2)!}_{=(2m+2)\cdot(2m+1)!}\\
&=\frac{1}{(m+1)\cdot m!}\cdot (1!\cdot 2!\cdot \cdots \cdot (2m)!)\cdot (2m+1)!\cdot (2m+2)\cdot (2m+1)!\\
&=\underbrace{\frac{2m+2}{m+1}}_{=2}\cdot \underbrace{\frac{1!\cdot 2!\cdot \cdots \cdot (2m)!}{m!}}_{=2^m\cdot\left(\prod_{i=1}^{m}((2i-1)!)\right)^2}\cdot ((2m+1)!)^2\\
&=2\cdot 2^m\cdot \left(\prod_{i=1}^{m}((2i-1)!)\right)^2\cdot ((2m+1)!)^2 .
\end{aligned}$$

Comparing this with

$$\begin{aligned}
&\underbrace{2^{m+1}}_{=2\cdot 2^m}\cdot \left(\underbrace{\prod_{i=1}^{m+1}((2i-1)!)}_{=\left(\prod_{i=1}^{m}((2i-1)!)\right)\cdot(2(m+1)-1)!}\right)^2\\
&=2\cdot 2^m\cdot \left(\left(\prod_{i=1}^{m}((2i-1)!)\right)\cdot \left(\underbrace{2(m+1)-1}_{=2m+1}\right)!\right)^2\\
&=2\cdot 2^m\cdot \left(\left(\prod_{i=1}^{m}((2i-1)!)\right)\cdot (2m+1)!\right)^2\\
&=2\cdot 2^m\cdot \left(\prod_{i=1}^{m}((2i-1)!)\right)^2\cdot ((2m+1)!)^2 ,
\end{aligned}$$

we obtain

$$\frac{1! \cdot 2! \cdot \cdots \cdot (2(m+1))!}{(m+1)!} = 2^{m+1} \cdot \left(\prod_{i=1}^{m+1} ((2i-1)!) \right)^2 .$$

In other words, Exercise 3.5 **(c)** holds for $n = m+1$. This completes the induction step. Thus, Exercise 3.5 **(c)** is solved. □

7.16. Solution to Exercise 3.6

In order to solve Exercise 3.6, we need to prove Proposition 3.21.

Proof of Proposition 3.21. We shall prove Proposition 3.21 by induction over n:

Induction base: We have

$$\sum_{k=0}^{0} \binom{0}{k} x^k y^{0-k} = \underbrace{\binom{0}{0}}_{\substack{=1 \\ \text{(by (227), applied to } m=0)}} \underbrace{x^0}_{=1} \underbrace{y^{0-0}}_{=y^0=1} = 1.$$

Comparing this with $(x+y)^0 = 1$, we obtain $(x+y)^0 = \sum_{k=0}^{0} \binom{0}{k} x^k y^{0-k}$. In other words, Proposition 3.21 holds for $n = 0$. This completes the induction base.

Induction step: Let N be a positive integer. Assume that Proposition 3.21 holds for $n = N-1$. We must now prove that Proposition 3.21 holds for $n = N$.

Notice that $N \geq 1$ (since N is a positive integer), so that $N - 1 \geq 0$.

We have assumed that Proposition 3.21 holds for $n = N-1$. In other words, we have

$$(x+y)^{N-1} = \sum_{k=0}^{N-1} \binom{N-1}{k} x^k y^{(N-1)-k}. \qquad (554)$$

Now,

$$\begin{aligned}
&(x+y)^N \\
&= (x+y) \underbrace{(x+y)^{N-1}}_{\substack{=\sum_{k=0}^{N-1}\binom{N-1}{k}x^k y^{(N-1)-k} \\ \text{(by (554))}}} \\
&= (x+y)\left(\sum_{k=0}^{N-1}\binom{N-1}{k}x^k y^{(N-1)-k}\right) \\
&= \underbrace{x\sum_{k=0}^{N-1}\binom{N-1}{k}x^k y^{(N-1)-k}}_{=\sum_{k=0}^{N-1}\binom{N-1}{k}xx^k y^{(N-1)-k}} + \underbrace{y\sum_{k=0}^{N-1}\binom{N-1}{k}x^k y^{(N-1)-k}}_{=\sum_{k=0}^{N-1}\binom{N-1}{k}x^k y y^{(N-1)-k}} \\
&= \sum_{k=0}^{N-1} \underbrace{\binom{N-1}{k}}_{\substack{=\binom{N-1}{(k+1)-1} \\ \text{(since } k=(k+1)-1)}} \underbrace{xx^k}_{=x^{k+1}} \underbrace{y^{(N-1)-k}}_{\substack{=y^{N-(k+1)} \\ \text{(since } (N-1)-k=N-(k+1))}} \\
&\qquad + \sum_{k=0}^{N-1}\binom{N-1}{k}x^k \underbrace{yy^{(N-1)-k}}_{\substack{=y^{((N-1)-k)+1}=y^{N-k} \\ \text{(since } ((N-1)-k)+1=N-k)}} \\
&= \underbrace{\sum_{k=0}^{N-1}\binom{N-1}{(k+1)-1}x^{k+1}y^{N-(k+1)}}_{\substack{=\sum_{k=1}^{N}\binom{N-1}{k-1}x^k y^{N-k} \\ \text{(here, we have substituted } k \text{ for } k+1 \text{ in the sum)}}} + \sum_{k=0}^{N-1}\binom{N-1}{k}x^k y^{N-k} \\
&= \sum_{k=1}^{N}\binom{N-1}{k-1}x^k y^{N-k} + \sum_{k=0}^{N-1}\binom{N-1}{k}x^k y^{N-k}.
\end{aligned} \tag{555}$$

But recall that $N-1 \geq 0$, so that $N-1 \in \mathbb{N}$. Clearly, $N-1 < N$. Hence, (231) (applied to $m = N-1$ and $n = N$) yields $\binom{N-1}{N} = 0$.

On the other hand, N is a positive integer. Hence, $N \in \{1, 2, \ldots, N\}$. Thus, we can split off the addend for $k = N$ from the sum $\sum_{k=1}^{N}\binom{N-1}{k}x^k y^{N-k}$. We thus

obtain

$$\sum_{k=1}^{N}\binom{N-1}{k}x^{k}y^{N-k}=\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}+\underbrace{\binom{N-1}{N}}_{=0}x^{N}y^{N-N}$$
$$=\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}+\underbrace{0x^{N}y^{N-N}}_{=0}$$
$$=\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}. \qquad (556)$$

Furthermore, $N-1\geq 0$, so that $0\in\{0,1,\ldots,N-1\}$. Hence, we can split off the addend for $k=0$ from the sum $\sum_{k=0}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}$. We thus obtain

$$\sum_{k=0}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}$$
$$=\underbrace{\binom{N-1}{0}}_{\substack{=1\\ \text{(by (227), applied to } m=N-1)}}x^{0}y^{N-0}+\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}$$
$$=x^{0}y^{N-0}+\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}.$$

Comparing this with

$$\underbrace{\binom{N}{0}}_{\substack{=1\\ \text{(by (227), applied to } m=N)}}x^{0}y^{N-0}+\underbrace{\sum_{k=1}^{N}\binom{N-1}{k}x^{k}y^{N-k}}_{\substack{=\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}\\ \text{(by (556))}}}$$
$$=x^{0}y^{N-0}+\sum_{k=1}^{N-1}\binom{N-1}{k}x^{k}y^{N-k},$$

we obtain

$$\sum_{k=0}^{N-1}\binom{N-1}{k}x^{k}y^{N-k}=\binom{N}{0}x^{0}y^{N-0}+\sum_{k=1}^{N}\binom{N-1}{k}x^{k}y^{N-k}. \qquad (557)$$

Now, (555) becomes

$$\begin{aligned}
&(x+y)^N \\
&= \sum_{k=1}^{N} \binom{N-1}{k-1} x^k y^{N-k} + \underbrace{\sum_{k=0}^{N-1} \binom{N-1}{k} x^k y^{N-k}}_{\substack{=\binom{N}{0} x^0 y^{N-0} + \sum_{k=1}^{N} \binom{N-1}{k} x^k y^{N-k} \\ \text{(by (557))}}} \\
&= \sum_{k=1}^{N} \binom{N-1}{k-1} x^k y^{N-k} + \binom{N}{0} x^0 y^{N-0} + \sum_{k=1}^{N} \binom{N-1}{k} x^k y^{N-k} \\
&= \binom{N}{0} x^0 y^{N-0} + \underbrace{\sum_{k=1}^{N} \binom{N-1}{k-1} x^k y^{N-k} + \sum_{k=1}^{N} \binom{N-1}{k} x^k y^{N-k}}_{=\sum_{k=1}^{N} \left(\binom{N-1}{k-1} + \binom{N-1}{k} \right) x^k y^{N-k}} \\
&= \binom{N}{0} x^0 y^{N-0} + \sum_{k=1}^{N} \left(\binom{N-1}{k-1} + \binom{N-1}{k} \right) x^k y^{N-k}.
\end{aligned}$$

Comparing this with

$$\begin{aligned}
&\sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k} \\
&= \binom{N}{0} x^0 y^{N-0} + \sum_{k=1}^{N} \underbrace{\binom{N}{k}}_{\substack{=\binom{N-1}{k-1} + \binom{N-1}{k} \\ \text{(by (234), applied to } m=N \text{ and } n=k)}} x^k y^{N-k} \\
&\qquad \text{(here, we have split off the addend for } k=0 \text{ from the sum)} \\
&= \binom{N}{0} x^0 y^{N-0} + \sum_{k=1}^{N} \left(\binom{N-1}{k-1} + \binom{N-1}{k} \right) x^k y^{N-k},
\end{aligned}$$

we obtain $(x+y)^N = \sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k}$. In other words, Proposition 3.21 holds for $n = N$. Hence, Proposition 3.21 is proven by induction. □

7.17. Solution to Exercise 3.7

Exercise 3.7 may look scary, but it is a straightforward exercise on induction (on b). To make our life a little bit easier, we shall slightly relax the condition $a \leq b$ to $b \geq a - 1$ (so that we can use the case $b = a - 1$ instead of $b = a$ as an induction base):

Proposition 7.13. Let k be a positive integer. Let a be a positive integer such that $k \leq a$. Let $b \in \{a-1, a, a+1, \ldots\}$. Then,

$$\frac{k-1}{k} \sum_{n=a}^{b} \frac{1}{\dbinom{n}{k}} = \frac{1}{\dbinom{a-1}{k-1}} - \frac{1}{\dbinom{b}{k-1}}.$$

(In particular, all fractions appearing in this equality are well-defined.)

Before we prove this proposition (which rather obviously encompasses the claim of Exercise 3.7), let us show a few lemmas:

Lemma 7.14. Let $m \in \mathbb{Q}$ and $n \in \{1, 2, 3, \ldots\}$. Then,

$$\binom{m}{n} = \frac{m-n+1}{n} \binom{m}{n-1}.$$

Proof of Lemma 7.14. We have $n \neq 0$ (since $n \in \{1, 2, 3, \ldots\}$). Hence, the fraction $\frac{m-n+1}{n}$ is well-defined.

We have $n \in \{1, 2, 3, \ldots\}$; in other words, n is a positive integer. Hence, $n! = n \cdot (n-1)!$.

We have $n \in \{1, 2, 3, \ldots\} \subseteq \mathbb{N}$. Thus, the definition of $\binom{m}{n}$ yields

$$\begin{aligned}
\binom{m}{n} &= \frac{m(m-1)\cdots(m-n+1)}{n!} \\
&= \frac{m(m-1)\cdots(m-n+1)}{n\cdot(n-1)!} \qquad \left(\text{since } n! = n\cdot(n-1)!\right) \\
&= \frac{1}{n\cdot(n-1)!} \cdot \underbrace{(m(m-1)\cdots(m-n+1))}_{\substack{=(m(m-1)\cdots(m-n+2))\cdot(m-n+1) \\ \text{(since } n \text{ is a positive integer)}}} \\
&= \frac{1}{n\cdot(n-1)!} \cdot (m(m-1)\cdots(m-n+2)) \cdot (m-n+1) \\
&= \frac{m-n+1}{n} \cdot \frac{m(m-1)\cdots(m-n+2)}{(n-1)!}.
\end{aligned} \tag{558}$$

Moreover, $n-1 \in \mathbb{N}$ (since $n \in \{1, 2, 3, \ldots\}$), so that the definition of $\binom{m}{n-1}$ yields

$$\binom{m}{n-1} = \frac{m(m-1)\cdots(m-(n-1)+1)}{(n-1)!} = \frac{m(m-1)\cdots(m-n+2)}{(n-1)!}$$
$$\left(\text{since } m-(n-1)+1 = m-n+2\right).$$

Multiplying this equality by $\dfrac{m-n+1}{n}$, we obtain

$$\frac{m-n+1}{n}\binom{m}{n-1} = \frac{m-n+1}{n} \cdot \frac{m(m-1)\cdots(m-n+2)}{(n-1)!}.$$

Comparing this with (558), we obtain $\dbinom{m}{n} = \dfrac{m-n+1}{n}\dbinom{m}{n-1}$. This proves Lemma 7.14. □

Lemma 7.15. Let k be a positive integer. Let a be a positive integer such that $k \leq a$. Let $b \in \{a-1, a, a+1, \ldots\}$.

(a) The fractions $\dfrac{k-1}{k}$, $\dfrac{1}{\dbinom{a-1}{k-1}}$ and $\dfrac{1}{\dbinom{b}{k-1}}$ are well-defined.

(b) For each $n \in \{a, a+1, \ldots, b\}$, the fraction $\dfrac{1}{\dbinom{n}{k}}$ is well-defined.

Proof of Lemma 7.15. We first observe that

$$\binom{m}{n} \neq 0 \tag{559}$$

for any $m \in \mathbb{N}$ and $n \in \mathbb{N}$ satisfying $m \geq n$.

[*Proof of (559):* Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$ be such that $m \geq n$. Thus, Proposition 3.4 yields $\dbinom{m}{n} = \dfrac{m!}{n!(m-n)!} \neq 0$ (since $m! \neq 0$ (since $m!$ is a positive integer)). This proves (559).]

We have $k \neq 0$ (since k is a positive integer); thus, the fraction $\dfrac{k-1}{k}$ is well-defined.

Also, $a-1 \in \mathbb{N}$ (since a is a positive integer) and $k-1 \in \mathbb{N}$ (since k is a positive integer) and $a-1 \geq k-1$ (since $\underbrace{k}_{\leq a} -1 \leq a-1$). Thus, (559) (applied to $a-1$ and $k-1$ instead of m and n) yields $\dbinom{a-1}{k-1} \neq 0$. Hence, the fraction $\dfrac{1}{\dbinom{a-1}{k-1}}$ is well-defined.

Also, $b \in \{a-1, a, a+1, \ldots\}$, so that $b \geq a-1 \geq k-1 \geq 0$ (since $k-1 \in \mathbb{N}$). Hence, $b \in \mathbb{N}$ (since $b \geq 0$ and $b \in \{a-1, a, a+1, \ldots\} \subseteq \mathbb{Z}$) and $k-1 \in \mathbb{N}$ and $b \geq k-1$. Thus, (559) (applied to b and $k-1$ instead of m and n) yields $\dbinom{b}{k-1} \neq 0$. Hence, the fraction $\dfrac{1}{\dbinom{b}{k-1}}$ is well-defined.

We have now shown that the fractions $\dfrac{k-1}{k}$, $\dfrac{1}{\dbinom{a-1}{k-1}}$ and $\dfrac{1}{\dbinom{b}{k-1}}$ are well-defined. This proves Lemma 7.15 **(a)**.

(b) Let $n \in \{a, a+1, \ldots, b\}$. Thus, $n \geq a > a-1 \geq 0$, so that $n \in \mathbb{N}$ (since $n \in \{a, a+1, \ldots, b\} \subseteq \mathbb{Z}$). Also, $k \in \mathbb{N}$. Furthermore, $n \geq a \geq k$ (since $k \leq a$). Hence, (559) (applied to n and k instead of m and n) yields $\dbinom{n}{k} \neq 0$. Hence, the fraction $\dfrac{1}{\dbinom{n}{k}}$ is well-defined. This proves Lemma 7.15 **(b)**. □

Proof of Proposition 7.13. All fractions appearing in Proposition 7.13 are well-defined (because of Lemma 7.15). It thus remains to prove that

$$\frac{k-1}{k} \sum_{n=a}^{b} \frac{1}{\dbinom{n}{k}} = \frac{1}{\dbinom{a-1}{k-1}} - \frac{1}{\dbinom{b}{k-1}}. \tag{560}$$

Forget that we fixed b. We thus need to prove that (560) holds for every $b \in \{a-1, a, a+1, \ldots\}$. We shall prove this by induction on b:

Induction base: Comparing

$$\frac{k-1}{k} \underbrace{\sum_{n=a}^{a-1} \frac{1}{\dbinom{n}{k}}}_{=(\text{empty sum})=0} = \frac{k-1}{k} \cdot 0 = 0$$

with $\dfrac{1}{\dbinom{a-1}{k-1}} - \dfrac{1}{\dbinom{a-1}{k-1}} = 0$, we conclude that $\dfrac{k-1}{k} \sum\limits_{n=a}^{a-1} \dfrac{1}{\dbinom{n}{k}} = \dfrac{1}{\dbinom{a-1}{k-1}} - \dfrac{1}{\dbinom{a-1}{k-1}}$. In other words, (560) holds for $b = a-1$. This completes the induction base.

Induction step: Let $\beta \in \{a, a+1, a+2, \ldots\}$. Assume that (560) holds for $b = \beta - 1$. We must prove that (560) holds for $b = \beta$.

We have assumed that (560) holds for $b = \beta - 1$. In other words, we have

$$\frac{k-1}{k} \sum_{n=a}^{\beta-1} \frac{1}{\dbinom{n}{k}} = \frac{1}{\dbinom{a-1}{k-1}} - \frac{1}{\dbinom{\beta-1}{k-1}}. \tag{561}$$

(In particular, all fractions appearing in this equality are well-defined.)

We have $\beta \geq a$ (since $\beta \in \{a, a+1, a+2, \ldots\}$).

We have $k \in \{1, 2, 3, \ldots\}$ (since k is a positive integer). Hence, Proposition 3.22 (applied to β and k instead of m and n) yields $\dbinom{\beta}{k} = \dfrac{\beta}{k}\dbinom{\beta-1}{k-1}$. Multiplying this equality by k, we obtain

$$k\binom{\beta}{k} = k \cdot \frac{\beta}{k}\binom{\beta-1}{k-1} = \beta\binom{\beta-1}{k-1}. \tag{562}$$

Now,

$$\begin{aligned}
&\frac{k-1}{k} \underbrace{\sum_{n=a}^{\beta} \frac{1}{\binom{n}{k}}}_{\substack{=\sum_{n=a}^{\beta-1} \frac{1}{\binom{n}{k}} + \frac{1}{\binom{\beta}{k}} \\ \text{(since } \beta \geq a\text{)}}} \\
&= \frac{k-1}{k}\left(\sum_{n=a}^{\beta-1} \frac{1}{\binom{n}{k}} + \frac{1}{\binom{\beta}{k}}\right) \\
&= \underbrace{\frac{k-1}{k}\sum_{n=a}^{\beta-1} \frac{1}{\binom{n}{k}}}_{\substack{=\frac{1}{\binom{a-1}{k-1}} - \frac{1}{\binom{\beta-1}{k-1}} \\ \text{(by (561))}}} + \underbrace{\frac{k-1}{k} \cdot \frac{1}{\binom{\beta}{k}}}_{\substack{=\frac{k-1}{k\binom{\beta}{k}} = \frac{k-1}{\beta\binom{\beta-1}{k-1}} \\ \text{(by (562))}}} \\
&= \frac{1}{\binom{a-1}{k-1}} - \frac{1}{\binom{\beta-1}{k-1}} + \frac{k-1}{\beta\binom{\beta-1}{k-1}} = \frac{1}{\binom{a-1}{k-1}} - \underbrace{\left(\frac{1}{\binom{\beta-1}{k-1}} - \frac{k-1}{\beta\binom{\beta-1}{k-1}}\right)}_{\substack{=\frac{\beta-(k-1)}{\beta\binom{\beta-1}{k-1}} = \frac{\beta-k+1}{\beta\binom{\beta-1}{k-1}} \\ \text{(since } \beta-(k-1)=\beta-k+1\text{)}}} \\
&= \frac{1}{\binom{a-1}{k-1}} - \frac{\beta-k+1}{\beta\binom{\beta-1}{k-1}}.
\end{aligned} \tag{563}$$

Also, Lemma 7.14 (applied to $m = \beta$ and $n = k$) yields $\dbinom{\beta}{k} = \dfrac{\beta-k+1}{k}\dbinom{\beta}{k-1}$.

Multiplying this equality by k, we obtain

$$k\binom{\beta}{k} = k \cdot \frac{\beta - k + 1}{k}\binom{\beta}{k-1} = (\beta - k + 1)\binom{\beta}{k-1}.$$

Comparing this with (562), we obtain

$$\beta\binom{\beta-1}{k-1} = (\beta - k + 1)\binom{\beta}{k-1}.$$

Therefore,

$$\frac{\beta - k + 1}{\beta\binom{\beta-1}{k-1}} = \frac{\beta - k + 1}{(\beta - k + 1)\binom{\beta}{k-1}} = \frac{1}{\binom{\beta}{k-1}}.$$

Hence, (563) becomes

$$\frac{k-1}{k}\sum_{n=a}^{\beta}\frac{1}{\binom{n}{k}} = \frac{1}{\binom{a-1}{k-1}} - \underbrace{\frac{\beta - k + 1}{\beta\binom{\beta-1}{k-1}}}_{=\frac{1}{\binom{\beta}{k-1}}} = \frac{1}{\binom{a-1}{k-1}} - \frac{1}{\binom{\beta}{k-1}}.$$

In other words, (560) holds for $b = \beta$. This completes the induction step. Thus, (560) is proven by induction. This completes the proof of Proposition 7.13. □

Solution to Exercise 3.7. From $b \geq a \geq a - 1$, we obtain $b \in \{a-1, a, a+1, \ldots\}$ (since b is an integer). Thus, Proposition 7.13 yields $\frac{k-1}{k}\sum_{n=a}^{b}\frac{1}{\binom{n}{k}} = \frac{1}{\binom{a-1}{k-1}} - \frac{1}{\binom{b}{k-1}}$. This solves Exercise 3.7. □

7.18. Solution to Exercise 3.8

Solution to Exercise 3.8. For every $N \in \mathbb{N}$, we let $[N]$ denote the N-element set $\{1, 2, \ldots, N\}$.

For every $i \in \mathbb{N}$ and $j \in \mathbb{N}$, we define a *filled (i,j)-set* to mean a subset S of $[i] \times [j]$ satisfying the following two conditions:

1. For every $k \in [i]$, at least one element of S has its first coordinate[305] equal to k.

[305] The *coordinates* of a pair (u,v) mean the entries u and v. Thus, the first coordinate of (u,v) is u.

2. For every $\ell \in [j]$, at least one element of S has its second coordinate equal to ℓ.

We can visualize subsets S of $[i] \times [j]$ as selections of boxes in a rectangular table that has i rows and j columns[306]. For instance, for $i = 3$ and $j = 4$, we can represent the subset $S = \{(1,1), (1,3), (2,2), (3,1), (3,3), (3,4)\}$ of $[i] \times [j]$ as the selection

X		X	
	X		
X		X	X

(564)

(where the rows are labelled $1, 2, 3$ from top to bottom, the columns are labelled $1, 2, 3, 4$ from left to right, as in a matrix, and where the elements of S are marked with X'es). Condition 1 then says that every row contains at least one selected box (i.e., at least one X); and Condition 2 says that every column contains at least one selected box. Our example (564) satisfies these two conditions, but (for instance) the subset

X	X		X
	X		
X			X

does not (it fails Condition 2).

For every $i \in \mathbb{N}$ and $j \in \mathbb{N}$, we define a nonnegative integer $c_{i,j}$ as the number of all filled (i, j)-sets which have n elements[307]. We shall now show that (263) is satisfied.

Indeed, let us first prove that any $x \in \mathbb{N}$ and $y \in \mathbb{N}$ satisfy

$$\binom{xy}{n} = \sum_{i=0}^{n} \sum_{j=0}^{n} c_{i,j} \binom{x}{i} \binom{y}{j}. \tag{565}$$

Keep in mind that (565) and (263) are different claims: The x and y in (565) are nonnegative integers, while the X and Y in (263) are indeterminates!

[*Proof of (565):* Let $x \in \mathbb{N}$ and $y \in \mathbb{N}$. Recall that $\binom{xy}{n}$ is the number of all n-element subsets of a given xy-element set. Since $[x] \times [y]$ is an xy-element set, we thus conclude that $\binom{xy}{n}$ is the number of all n-element subsets of $[x] \times [y]$.

Now, let us find a different way to count all n-element subsets of $[x] \times [y]$. As above, we can visualize such subsets as selections of boxes in a rectangular table that has x rows and y columns; we again mark the selected boxes by X'es. We want

[306] Namely, for every $(u, v) \in S$, we select the box in row u and column v.

[307] When we say "have n elements", we mean "have exactly n elements", not "have at least n elements".

to count all ways to select n boxes in this table, i.e., to place n X'es in the table. We can place n X'es in the table by means of the following process:

1. We choose how many rows of the table will have at least one X. This can be a number from 0 to n (inclusive)[308]; we denote it by i.

2. We choose how many columns of the table will have at least one X. This can be a number from 0 to n (inclusive)[309]; we denote it by j.

3. We choose the i rows of the table that will have at least one X. This can be done in $\dbinom{x}{i}$ ways (since there are x rows to choose from).

4. We choose the j columns of the table that will have at least one X. This can be done in $\dbinom{y}{j}$ ways (since there are y columns to choose from).

5. It remains to place n X'es in the table in such a way that the rows that contain at least one X are precisely the i chosen rows, and the columns that contain at least one X are precisely the j chosen columns. To do so, we can temporarily remove all the remaining $x-i$ rows and $y-j$ columns. We are then left with a rectangular table that has i rows and j columns, and now we need to place n X'es in it in such a way that every row contains at least one X and every column contains at least one X. As we know, the number of ways to do this is $c_{i,j}$ (because this is how $c_{i,j}$ was defined).

This process makes it clear that the total number of ways to place n X'es in the (original) table is $\sum\limits_{i=0}^{n}\sum\limits_{j=0}^{n}\dbinom{x}{i}\dbinom{y}{j}c_{i,j} = \sum\limits_{i=0}^{n}\sum\limits_{j=0}^{n}c_{i,j}\dbinom{x}{i}\dbinom{y}{j}$. In other words, the number of all n-element subsets of $[x]\times[y]$ is $\sum\limits_{i=0}^{n}\sum\limits_{j=0}^{n}c_{i,j}\dbinom{x}{i}\dbinom{y}{j}$.

So we know that this number equals both $\dbinom{xy}{n}$ and $\sum\limits_{i=0}^{n}\sum\limits_{j=0}^{n}c_{i,j}\dbinom{x}{i}\dbinom{y}{j}$ at the same time. Comparing these values, we obtain (565).]

Now that (565) is proven, we can finally solve the exercise. We define two polynomials P and Q in the indeterminates X and Y with rational coefficients by setting

$$P = \binom{XY}{n};$$
$$Q = \sum_{i=0}^{n}\sum_{j=0}^{n}c_{i,j}\binom{X}{i}\binom{Y}{j}$$

[308] I am not saying that any number from 0 to n (inclusive) is possible; I am just saying that this will always be a number from 0 to n (inclusive). Here is why:

Clearly, the number of rows of the table that will have at least one X is ≥ 0. But it is also $\leq n$, because we want to place only n X'es in the table, and these n X'es will clearly occupy at most n rows.

[309] This follows by a similar argument as the analogous statement in Step 1.

[310]. The equality (565) (which we have proven) states that $P(x,y) = Q(x,y)$ for all $x \in \mathbb{N}$ and $y \in \mathbb{N}$. Thus, Lemma 3.28 **(d)** yields that $P = Q$. Recalling how P and Q are defined, we see that this is precisely the equality (263).

Hence, (263) is proven, and Exercise 3.8 solved. □

Remark 7.16. I learnt the above solution to Exercise 3.8 from Gjergji Zaimi on AoPS. The numbers $c_{i,j}$ constructed in the solution do not appear to be easily computable by a simple closed formula. Nevertheless, they have some nice properties that can be easily obtained from their combinatorial definition:

- We have $c_{i,j} = c_{j,i}$ for all $i \in \mathbb{N}$ and $j \in \mathbb{N}$.
- We have $c_{0,0} = \begin{cases} 1, & \text{if } n = 0; \\ 0, & \text{if } n > 0 \end{cases}$, but every positive integer j satisfies $c_{0,j} = 0$.
- We have $c_{1,j} = \begin{cases} 1, & \text{if } n = j; \\ 0, & \text{if } n \neq j \end{cases}$ for all positive integers j.
- We have $c_{2,2} = \begin{cases} 0, & \text{if } n \leq 1; \\ 2, & \text{if } n = 2; \\ 4, & \text{if } n = 3; \\ 1, & \text{if } n = 4; \\ 0, & \text{if } n > 4 \end{cases}$.
- We have $c_{i,j} = 0$ if $n > ij$.
- For every $j \in \mathbb{N}$, the number $c_{n,j}$ is the number of all surjective maps $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, j\}$.

7.19. Solution to Exercise 3.9

Solution to Exercise 3.9. Here is one possible solution:

Exercise 3.8 shows that, for every $n \in \mathbb{N}$, there exist **nonnegative** integers $c_{i,j}$ for all $0 \leq i \leq n$ and $0 \leq j \leq n$ such that (263) holds. We denote these integers $c_{i,j}$ by $c_{i,j,n}$ (in order to make their dependence on n explicit). Thus, for every $n \in \mathbb{N}$, the nonnegative integers $c_{i,j,n}$ defined for all $0 \leq i \leq n$ and $0 \leq j \leq n$ satisfy

$$\binom{XY}{n} = \sum_{i=0}^{n} \sum_{j=0}^{n} c_{i,j,n} \binom{X}{i} \binom{Y}{j}. \tag{566}$$

[310]These are both polynomials since $\binom{XY}{n}$, $\binom{X}{i}$ and $\binom{Y}{j}$ are polynomials in X and Y.

Substituting a and X for X and Y in this equality, we obtain

$$\binom{aX}{n} = \sum_{i=0}^{n} \sum_{j=0}^{n} c_{i,j,n} \binom{a}{i} \binom{X}{j}. \qquad (567)$$

Now, Theorem 3.30 (applied to $n = c$) yields

$$\binom{X+Y}{c} = \sum_{k=0}^{c} \binom{X}{k} \binom{Y}{c-k}.$$

Substituting aX and b for X and Y in this equality, we obtain

$$\begin{aligned}
\binom{aX+b}{c} &= \sum_{k=0}^{c} \underbrace{\binom{aX}{k}}_{\substack{=\sum_{i=0}^{k}\sum_{j=0}^{k} c_{i,j,k}\binom{a}{i}\binom{X}{j} \\ \text{(by (567), applied to } n=k)}} \binom{b}{c-k} \\
&= \sum_{k=0}^{c} \left(\sum_{i=0}^{k} \sum_{j=0}^{k} c_{i,j,k} \binom{a}{i} \binom{X}{j} \right) \binom{b}{c-k} \\
&= \underbrace{\sum_{k=0}^{c} \sum_{i=0}^{k} \sum_{j=0}^{k}}_{=\sum_{j=0}^{c}\sum_{k=j}^{c}\sum_{i=0}^{k}} c_{i,j,k} \binom{a}{i} \underbrace{\binom{X}{j} \binom{b}{c-k}}_{=\binom{b}{c-k}\binom{X}{j}} \\
&= \sum_{j=0}^{c} \sum_{k=j}^{c} \sum_{i=0}^{k} c_{i,j,k} \binom{a}{i} \binom{b}{c-k} \binom{X}{j}. \qquad (568)
\end{aligned}$$

Now, for every $j \in \{0, 1, \ldots, c\}$, define an integer d_j by $d_j = \sum_{k=j}^{c} \sum_{i=0}^{k} c_{i,j,k} \binom{a}{i} \binom{b}{c-k}$. This d_j is clearly a nonnegative integer (since the $c_{i,j,n}$ are nonnegative, and so are the binomial coefficients $\binom{a}{i}$ and $\binom{b}{c-k}$ (due to a and b being nonnegative)). Then, (568) becomes

$$\binom{aX+b}{c} = \sum_{j=0}^{c} \underbrace{\sum_{k=j}^{c} \sum_{i=0}^{k} c_{i,j,k} \binom{a}{i} \binom{b}{c-k}}_{=d_j} \binom{X}{j} = \sum_{j=0}^{c} d_j \binom{X}{j} = \sum_{i=0}^{c} d_i \binom{X}{i}.$$

Exercise 3.9 is thus solved. □

7.20. Solution to Exercise 3.10

Before we come to the solution of Exercise 3.10, let us show a simple identity:

Proposition 7.17. Let $m \in \mathbb{N}$ and $p \in \mathbb{N}$ be such that $p \geq m$. Then,

$$\sum_{k=0}^{p} \binom{m}{k} = 2^m.$$

Proof of Proposition 7.17. For each $k \in \{m+1, m+2, \ldots, p\}$, we have

$$\binom{m}{k} = 0 \tag{569}$$

[311].

We have $p \geq m \geq 0$ (since $m \in \mathbb{N}$). Thus, the sum $\sum_{k=0}^{p} \binom{m}{k}$ can be split as follows:

$$\sum_{k=0}^{p} \binom{m}{k} = \sum_{k=0}^{m} \binom{m}{k} + \sum_{k=m+1}^{p} \underbrace{\binom{m}{k}}_{\substack{=0 \\ \text{(by (569))}}} = \sum_{k=0}^{m} \binom{m}{k} + \underbrace{\sum_{k=m+1}^{p} 0}_{=0} = \sum_{k=0}^{m} \binom{m}{k} = 2^m$$

(by Proposition 3.39 **(b)** (applied to $n = m$)). This proves Proposition 7.17. □

[311] *Proof of (569):* Let $k \in \{m+1, m+2, \ldots, p\}$. Thus, $k \geq m+1 > m$ and therefore $m < k$. Also, $k > m \geq 0$ (since $m \in \mathbb{N}$). Hence, $k \in \mathbb{N}$. Therefore, Proposition 3.6 (applied to $n = k$) yields $\binom{m}{k} = 0$. This proves (569).

Solution to Exercise 3.10. **(a)** For each $k \in \mathbb{N}$, we have

$$\underbrace{\dbinom{n}{k}}_{\substack{=\dfrac{n(n-1)\cdots(n-k+1)}{k!} \\ \text{(by (226) (applied to } n \text{ and } k \\ \text{instead of } m \text{ and } n))}} \underbrace{\dbinom{n-k}{b}}_{\substack{=\dfrac{(n-k)((n-k)-1)\cdots((n-k)-b+1)}{b!} \\ \text{(by (226) (applied to } n-k \text{ and } b \\ \text{instead of } m \text{ and } n))}}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \frac{(n-k)((n-k)-1)\cdots((n-k)-b+1)}{b!}$$

$$= \frac{1}{k!b!} \cdot (n(n-1)\cdots(n-k+1)) \cdot \left(\underbrace{(n-k)((n-k)-1)\cdots((n-k)-b+1)}_{=(n-k)(n-k-1)\cdots(n-k-b+1)} \right)$$

$$= \frac{1}{k!b!} \cdot \underbrace{(n(n-1)\cdots(n-k+1)) \cdot ((n-k)(n-k-1)\cdots(n-k-b+1))}_{=n(n-1)\cdots(n-k-b+1)}$$

$$= \frac{1}{k!b!} \cdot (n(n-1)\cdots(n-k-b+1)). \tag{570}$$

Let $j \geq a$ be an integer. Thus, $j \geq a \geq 0$ (since $a \in \mathbb{N}$), so that $j \in \mathbb{N}$. Hence, Proposition 3.4 (applied to j and a instead of m and n) shows that

$$\dbinom{j}{a} = \frac{j!}{a!(j-a)!}.$$

But (570) (applied to $k = j$) shows that

$$\dbinom{n}{j}\dbinom{n-j}{b} = \frac{1}{j!b!} \cdot (n(n-1)\cdots(n-j-b+1)).$$

Multiplying these two equalities, we obtain

$$\begin{aligned} \dbinom{j}{a}\dbinom{n}{j}\dbinom{n-j}{b} &= \frac{j!}{a!(j-a)!} \cdot \frac{1}{j!b!} \cdot (n(n-1)\cdots(n-j-b+1)) \\ &= \frac{1}{a!(j-a)!b!} \cdot (n(n-1)\cdots(n-j-b+1)). \end{aligned} \tag{571}$$

On the other hand, set $m = n - a - b$. Then,

$$\underbrace{m}_{=n-a-b} - (j-a) = (n-a-b) - (j-a) = n - j - b$$

and $n \geq m$ (since $m = n - \underbrace{a}_{\geq 0} - \underbrace{b}_{\geq 0} \leq n$) and $m \geq m - (j-a)$ (since $m - \left(\underbrace{j}_{\geq a} - a \right) \leq m - (a-a) = m$).

The equality (570) (applied to $k=a$) yields

$$\binom{n}{a}\binom{n-a}{b}=\frac{1}{a!b!}\cdot\left(n\left(n-1\right)\cdots\left(n-a-b+1\right)\right)=\frac{1}{a!b!}\cdot\left(n\left(n-1\right)\cdots\left(m+1\right)\right)$$

(since $n-a-b=m$). Furthermore, $j-a\in\mathbb{N}$ (since $j\geq a$), and thus the binomial coefficient $\binom{m}{j-a}$ is well-defined. Furthermore, from $n-a-b=m$, we obtain

$$\binom{n-a-b}{j-a}=\binom{m}{j-a}=\frac{m\left(m-1\right)\cdots\left(m-\left(j-a\right)+1\right)}{\left(j-a\right)!}$$

(by (226) (applied to $j-a$ instead of n)). Now,

$$\begin{aligned}
&\underbrace{\binom{n}{a}\binom{n-a}{b}}_{=\frac{1}{a!b!}\cdot\left(n\left(n-1\right)\cdots\left(m+1\right)\right)}\underbrace{\binom{n-a-b}{j-a}}_{=\frac{m\left(m-1\right)\cdots\left(m-\left(j-a\right)+1\right)}{\left(j-a\right)!}}\\
&=\frac{1}{a!b!}\cdot\left(n\left(n-1\right)\cdots\left(m+1\right)\right)\cdot\frac{m\left(m-1\right)\cdots\left(m-\left(j-a\right)+1\right)}{\left(j-a\right)!}\\
&=\frac{1}{a!\left(j-a\right)!b!}\cdot\underbrace{\left(n\left(n-1\right)\cdots\left(m+1\right)\right)\cdot\left(m\left(m-1\right)\cdots\left(m-\left(j-a\right)+1\right)\right)}_{\substack{=n\left(n-1\right)\cdots\left(m-\left(j-a\right)+1\right)\\ \text{(since } n\geq m\geq m-\left(j-a\right)\text{)}}}\\
&=\frac{1}{a!\left(j-a\right)!b!}\cdot\left(n\left(n-1\right)\cdots\left(m-\left(j-a\right)+1\right)\right)\\
&=\frac{1}{a!\left(j-a\right)!b!}\cdot\left(n\left(n-1\right)\cdots\left(n-j-b+1\right)\right)\qquad\left(\text{since } m-\left(j-a\right)=n-j-b\right).
\end{aligned}$$

Comparing this with (571), we obtain

$$\binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{j-a}=\binom{j}{a}\binom{n}{j}\binom{n-j}{b}=\binom{n}{j}\binom{j}{a}\binom{n-j}{b}.$$

This solves Exercise 3.10 **(a)**.

(b) Let $n\geq a$ be an integer. Thus, $n-a\in\mathbb{N}$. Now, we claim that

$$\sum_{j=a}^{n}\binom{n}{j}\binom{j}{a}\binom{n-j}{b}=\binom{n}{a}\binom{n-a}{b}2^{n-a-b}. \qquad (572)$$

[*Proof of (572):* First of all, (572) holds if $n-a-b<0$ [312]. Hence, for the rest of this proof, we can WLOG assume that we don't have $n-a-b<0$. Assume this.

We have $n-a-b\geq 0$ (since we don't have $n-a-b<0$), so that $n-a-b\in\mathbb{N}$. Thus, we have $n-a\in\mathbb{N}$ and $n-a-b\in\mathbb{N}$ and $n-a\geq n-a-b$ (since

[312] *Proof.* Assume that $n-a-b<0$. Thus, $n-a<b$. Hence, Proposition 3.6 (applied to $n-a$ and b

$(n-a)-(n-a-b)=b\geq 0$). Hence, Proposition 7.17 (applied to $m=n-a-b$ and $p=n-a$) yields

$$\sum_{k=0}^{n-a}\binom{n-a-b}{k}=2^{n-a-b}.$$

Now,

$$\begin{aligned}
&\sum_{j=a}^{n}\underbrace{\binom{n}{j}\binom{j}{a}\binom{n-j}{b}}_{\substack{=\binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{j-a}\\ \text{(by Exercise 3.10 (a))}}}\\
&=\sum_{j=a}^{n}\binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{j-a}=\binom{n}{a}\binom{n-a}{b}\sum_{j=a}^{n}\binom{n-a-b}{j-a}\\
&=\binom{n}{a}\binom{n-a}{b}\underbrace{\sum_{k=0}^{n-a}\binom{n-a-b}{k}}_{=2^{n-a-b}}\\
&\qquad\text{(here, we have substituted } k \text{ for } j-a \text{ in the sum)}\\
&=\binom{n}{a}\binom{n-a}{b}2^{n-a-b}.
\end{aligned}$$

This proves (572).]

Clearly, (572) answers Exercise 3.10 **(b)**. □

---

instead of m and n) shows that $\binom{n-a}{b}=0$ (since $n-a\in\mathbb{N}$). But

$$\begin{aligned}
\sum_{j=a}^{n}\underbrace{\binom{n}{j}\binom{j}{a}\binom{n-j}{b}}_{\substack{=\binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{j-a}\\ \text{(by Exercise 3.10 (a))}}}&=\sum_{j=a}^{n}\binom{n}{a}\underbrace{\binom{n-a}{b}}_{=0}\binom{n-a-b}{j-a}\\
&=\sum_{j=a}^{n}\binom{n}{a}0\binom{n-a-b}{j-a}=0.
\end{aligned}$$

Comparing this with

$$\binom{n}{a}\underbrace{\binom{n-a}{b}}_{=0}2^{n-a-b}=0,$$

we obtain $\sum_{j=a}^{n}\binom{n}{j}\binom{j}{a}\binom{n-j}{b}=\binom{n}{a}\binom{n-a}{b}2^{n-a-b}$. Hence, (572) is proven under the assumption that $n-a-b<0$.

7.21. Solution to Exercise 3.11

There are many ways to solve Exercise 3.11. The following one might be the shortest:

Proof of Lemma 3.47. We proceed by induction over m:

Induction base: We have

$$\sum_{r=0}^{k} (-1)^r \binom{n}{r} \binom{r}{k}$$
$$= \sum_{r=0}^{k-1} (-1)^r \binom{n}{r} \underbrace{\binom{r}{k}}_{\substack{=0 \\ \text{(by Proposition 3.6 (applied to } r \text{ and } k \\ \text{instead of } m \text{ and } n\text{) (since } r \leq k-1 < k\text{))}}} + (-1)^k \binom{n}{k} \underbrace{\binom{k}{k}}_{\substack{=1 \\ \text{(by Proposition 3.9} \\ \text{(applied to } k \text{ instead of } m\text{))}}}$$
$$\left(\text{here, we have split off the addend for } r = k \text{ from the sum}\right)$$
$$= \underbrace{\sum_{r=0}^{k-1} (-1)^r \binom{n}{r} 0}_{=0} + (-1)^k \binom{n}{k} = (-1)^k \binom{n}{k}.$$

Comparing this with

$$(-1)^k \binom{n}{k} \underbrace{\binom{n-k-1}{k-k}}_{\substack{= \binom{n-k-1}{0} = 1 \\ \text{(by Proposition 3.3 (a)} \\ \text{(applied to } n-k-1 \text{ instead of } m\text{))}}} = (-1)^k \binom{n}{k},$$

we obtain $\sum_{r=0}^{k} (-1)^r \binom{n}{r} \binom{r}{k} = (-1)^k \binom{n}{k} \binom{n-k-1}{k-k}$. In other words, Lemma 3.47 holds for $m = k$. This completes the induction base.

Induction step: Let M be an element of $\{k, k+1, k+2, \ldots\}$ such that $M > k$. Assume that Lemma 3.47 holds for $m = M - 1$. We must now prove that Lemma 3.47 holds for $m = M$.

We have assumed that Lemma 3.47 holds for $m = M - 1$. In other words, we have

$$\sum_{r=0}^{M-1} (-1)^r \binom{n}{r} \binom{r}{k} = (-1)^{M-1} \binom{n}{k} \binom{n-k-1}{(M-1)-k}.$$

Hence,

$$\sum_{r=0}^{M-1} (-1)^r \binom{n}{r} \binom{r}{k} = \underbrace{(-1)^{M-1}}_{=-(-1)^M} \binom{n}{k} \underbrace{\binom{n-k-1}{(M-1)-k}}_{\substack{=\binom{n-k-1}{M-k-1} \\ (\text{since } (M-1)-k=M-k-1)}}$$
$$= -(-1)^M \binom{n}{k} \binom{n-k-1}{M-k-1}. \tag{573}$$

On the other hand, $M \geq k$ (since $M > k$). Hence, Proposition 3.23 (applied to n, M and k instead of m, i and a) yields

$$\binom{n}{M} \binom{M}{k} = \binom{n}{k} \binom{n-k}{M-k}. \tag{574}$$

Furthermore, $M - k > 0$ (since $M > k$), so that $M - k \in \{1, 2, 3, \ldots\}$ (since $M - k$ is an integer). Hence, Proposition 3.11 (applied to $n - k$ and $M - k$ instead of m and n) yields

$$\binom{n-k}{M-k} = \binom{n-k-1}{M-k-1} + \binom{n-k-1}{M-k}.$$

Hence, (574) becomes

$$\binom{n}{M} \binom{M}{k} = \binom{n}{k} \underbrace{\binom{n-k}{M-k}}_{=\binom{n-k-1}{M-k-1}+\binom{n-k-1}{M-k}}$$
$$= \binom{n}{k} \left(\binom{n-k-1}{M-k-1} + \binom{n-k-1}{M-k} \right). \tag{575}$$

Now,

$$\sum_{r=0}^{M}(-1)^r\binom{n}{r}\binom{r}{k}$$
$$= \underbrace{\sum_{r=0}^{M-1}(-1)^r\binom{n}{r}\binom{r}{k}}_{\substack{=-(-1)^M\binom{n}{k}\binom{n-k-1}{M-k-1} \\ \text{(by (573))}}} + (-1)^M \underbrace{\binom{n}{M}\binom{M}{k}}_{\substack{=\binom{n}{k}\left(\binom{n-k-1}{M-k-1}+\binom{n-k-1}{M-k}\right) \\ \text{(by (575))}}}$$

(here, we have split off the addend for $r = M$ from the sum)

$$= -(-1)^M\binom{n}{k}\binom{n-k-1}{M-k-1} + \underbrace{(-1)^M\binom{n}{k}\left(\binom{n-k-1}{M-k-1}+\binom{n-k-1}{M-k}\right)}_{=(-1)^M\binom{n}{k}\binom{n-k-1}{M-k-1}+(-1)^M\binom{n}{k}\binom{n-k-1}{M-k}}$$
$$= -(-1)^M\binom{n}{k}\binom{n-k-1}{M-k-1} + (-1)^M\binom{n}{k}\binom{n-k-1}{M-k-1} + (-1)^M\binom{n}{k}\binom{n-k-1}{M-k}$$
$$= (-1)^M\binom{n}{k}\binom{n-k-1}{M-k}.$$

In other words, Lemma 3.47 holds for $m = M$. This completes the induction step. Hence, Lemma 3.47 is proven. □

We notice that Lemma 3.47 holds even if we replace the assumption "$n \in \mathbb{N}$" by "$n \in \mathbb{Q}$". In fact, our above proof of Lemma 3.47 applies verbatim in this more general setting.

Solution to Exercise 3.11. We have proven Lemma 3.47; hence, Exercise 3.11 is solved. □

7.22. Solution to Exercise 3.12

Solution to Exercise 3.12. **(a)** Let $\mathcal{A}$ and $\mathcal{B}$ be two equivalent statements.

We have $[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases}$ (by the definition of $[\mathcal{A}]$) and $[\mathcal{B}] = \begin{cases} 1, & \text{if } \mathcal{B} \text{ is true;} \\ 0, & \text{if } \mathcal{B} \text{ is false} \end{cases}$ (by the definition of $[\mathcal{B}]$).

But $\mathcal{A}$ and $\mathcal{B}$ are equivalent. Thus, $\mathcal{A}$ is true (resp. false) if and only if $\mathcal{B}$ is true (resp. false). Hence, $\begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases} = \begin{cases} 1, & \text{if } \mathcal{B} \text{ is true;} \\ 0, & \text{if } \mathcal{B} \text{ is false} \end{cases}$. Thus,

$$[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases} = \begin{cases} 1, & \text{if } \mathcal{B} \text{ is true;} \\ 0, & \text{if } \mathcal{B} \text{ is false} \end{cases} = [\mathcal{B}].$$

This solves Exercise 3.12 **(a)**.

(b) Let $\mathcal{A}$ be any logical statement. Then, (not $\mathcal{A}$) is true (resp. false) if and only if $\mathcal{A}$ is false (resp. true). Hence,

$$\begin{cases} 1, & \text{if } (\text{not } \mathcal{A}) \text{ is true;} \\ 0, & \text{if } (\text{not } \mathcal{A}) \text{ is false} \end{cases} = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is false;} \\ 0, & \text{if } \mathcal{A} \text{ is true} \end{cases} = \begin{cases} 0, & \text{if } \mathcal{A} \text{ is true;} \\ 1, & \text{if } \mathcal{A} \text{ is false} \end{cases}.$$

Now, the definition of $[\text{not } \mathcal{A}]$ shows that

$$[\text{not } \mathcal{A}] = \begin{cases} 1, & \text{if } (\text{not } \mathcal{A}) \text{ is true;} \\ 0, & \text{if } (\text{not } \mathcal{A}) \text{ is false} \end{cases} = \begin{cases} 0, & \text{if } \mathcal{A} \text{ is true;} \\ 1, & \text{if } \mathcal{A} \text{ is false} \end{cases}.$$

Adding this equality to

$$[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases},$$

we obtain

$$\begin{aligned} [\mathcal{A}] + [\text{not } \mathcal{A}] &= \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases} + \begin{cases} 0, & \text{if } \mathcal{A} \text{ is true;} \\ 1, & \text{if } \mathcal{A} \text{ is false} \end{cases} = \begin{cases} 1+0, & \text{if } \mathcal{A} \text{ is true;} \\ 0+1, & \text{if } \mathcal{A} \text{ is false} \end{cases} \\ &= \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 1, & \text{if } \mathcal{A} \text{ is false} \end{cases} = 1. \end{aligned}$$

Thus, $[\text{not } \mathcal{A}] = 1 - [\mathcal{A}]$. This solves Exercise 3.12 **(b)**.

(c) Let $\mathcal{A}$ and $\mathcal{B}$ be two logical statements. We must be in one of the following two cases:

Case 1: The statement $\mathcal{A}$ is true.

Case 2: The statement $\mathcal{A}$ is false.

Let us consider Case 1 first. In this case, the statement $\mathcal{A}$ is true. Hence, the statement $\mathcal{A} \wedge \mathcal{B}$ is equivalent to the statement $\mathcal{B}$. Thus, Exercise 3.12 **(a)** (applied to $\mathcal{A} \wedge \mathcal{B}$ instead of $\mathcal{A}$) shows that $[\mathcal{A} \wedge \mathcal{B}] = [\mathcal{B}]$. But the definition of $[\mathcal{A}]$ yields $[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases} = 1$ (since $\mathcal{A}$ is true). Hence, $\underbrace{[\mathcal{A}]}_{=1} [\mathcal{B}] = [\mathcal{B}]$. Comparing this with $[\mathcal{A} \wedge \mathcal{B}] = [\mathcal{B}]$, we obtain $[\mathcal{A} \wedge \mathcal{B}] = [\mathcal{A}] [\mathcal{B}]$. Thus, Exercise 3.12 **(c)** is solved in Case 1.

Let us now consider Case 2. In this case, the statement $\mathcal{A}$ is false. Hence, the statement $\mathcal{A} \wedge \mathcal{B}$ is false as well. Thus, the definition of $[\mathcal{A} \wedge \mathcal{B}]$ yields $[\mathcal{A} \wedge \mathcal{B}] = \begin{cases} 1, & \text{if } \mathcal{A} \wedge \mathcal{B} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \wedge \mathcal{B} \text{ is false} \end{cases} = 0$ (since $\mathcal{A} \wedge \mathcal{B}$ is false). But the definition of $[\mathcal{A}]$ yields $[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases} = 0$ (since $\mathcal{A}$ is false). Hence, $\underbrace{[\mathcal{A}]}_{=0} [\mathcal{B}] = 0 [\mathcal{B}] = 0$.

Comparing this with $[\mathcal{A} \wedge \mathcal{B}] = 0$, we obtain $[\mathcal{A} \wedge \mathcal{B}] = [\mathcal{A}] [\mathcal{B}]$. Thus, Exercise 3.12 **(c)** is solved in Case 2.

We thus have solved Exercise 3.12 **(c)** in both Cases 1 and 2. Hence, Exercise 3.12 **(c)** always holds.

[*Remark:* It is, of course, also possible to get a completely straightforward solution to Exercise 3.12 **(c)** by distinguishing four cases, depending on which of the statements $\mathcal{A}$ and $\mathcal{B}$ are true.]

(d) It is easy to solve Exercise 3.12 **(d)** by a case distinction similarly to Exercise 3.12 **(c)**. However, since we have already solved parts **(b)** and **(c)**, we can give a simpler solution:

Let $\mathcal{A}$ and $\mathcal{B}$ be two logical statements. One of de Morgan's laws says that the statement $(\text{not } (\mathcal{A} \vee \mathcal{B}))$ is equivalent to $(\text{not } \mathcal{A}) \wedge (\text{not } \mathcal{B})$. Hence, Exercise 3.12 **(a)** (applied to $(\text{not } (\mathcal{A} \vee \mathcal{B}))$ and $(\text{not } \mathcal{A}) \wedge (\text{not } \mathcal{B})$ instead of $\mathcal{A}$ and $\mathcal{B}$) shows that

$$\begin{aligned}[\text{not } (\mathcal{A} \vee \mathcal{B})] &= [(\text{not } \mathcal{A}) \wedge (\text{not } \mathcal{B})] \\ &= \underbrace{[\text{not } \mathcal{A}]}_{\substack{=1-[\mathcal{A}] \\ \text{(by Exercise 3.12 (b))}}} \underbrace{[\text{not } \mathcal{B}]}_{\substack{=1-[\mathcal{B}] \\ \text{(by Exercise 3.12 (b),} \\ \text{applied to } \mathcal{B} \text{ instead of } \mathcal{A})}} \\ &\qquad \left(\begin{array}{c}\text{by Exercise 3.12 (c), applied to} \\ (\text{not } \mathcal{A}) \text{ and } (\text{not } \mathcal{B}) \text{ instead of } \mathcal{A} \text{ and } \mathcal{B}\end{array}\right) \\ &= (1 - [\mathcal{A}]) (1 - [\mathcal{B}]) = 1 - [\mathcal{A}] - [\mathcal{B}] + [\mathcal{A}] [\mathcal{B}] .\end{aligned}$$

But Exercise 3.12 **(b)** (applied to $\mathcal{A} \vee \mathcal{B}$ instead of $\mathcal{A}$) shows that $[\text{not } (\mathcal{A} \vee \mathcal{B})] = 1 - [\mathcal{A} \vee \mathcal{B}]$. Hence,

$$[\mathcal{A} \vee \mathcal{B}] = 1 - \underbrace{[\text{not } (\mathcal{A} \vee \mathcal{B})]}_{=1-[\mathcal{A}]-[\mathcal{B}]+[\mathcal{A}][\mathcal{B}]} = 1 - (1 - [\mathcal{A}] - [\mathcal{B}] + [\mathcal{A}] [\mathcal{B}]) = [\mathcal{A}] + [\mathcal{B}] - [\mathcal{A}] [\mathcal{B}] .$$

This solves Exercise 3.12 **(d)**.

(e) Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be three logical statements. Then, Exercise 3.12 **(d)** (applied to $\mathcal{A} \vee \mathcal{B}$ and $\mathcal{C}$ instead of $\mathcal{A}$ and $\mathcal{B}$) shows that

$$\begin{aligned}[(\mathcal{A} \vee \mathcal{B}) \vee \mathcal{C}] &= \underbrace{[\mathcal{A} \vee \mathcal{B}]}_{\substack{=[\mathcal{A}]+[\mathcal{B}]-[\mathcal{A}][\mathcal{B}] \\ \text{(by Exercise 3.12 (d))}}} + [\mathcal{C}] - \underbrace{[\mathcal{A} \vee \mathcal{B}]}_{\substack{=[\mathcal{A}]+[\mathcal{B}]-[\mathcal{A}][\mathcal{B}] \\ \text{(by Exercise 3.12 (d))}}} [\mathcal{C}] \\ &= ([\mathcal{A}] + [\mathcal{B}] - [\mathcal{A}] [\mathcal{B}]) + [\mathcal{C}] - ([\mathcal{A}] + [\mathcal{B}] - [\mathcal{A}] [\mathcal{B}]) [\mathcal{C}] \\ &= [\mathcal{A}] + [\mathcal{B}] + [\mathcal{C}] - [\mathcal{A}] [\mathcal{B}] - [\mathcal{A}] [\mathcal{C}] - [\mathcal{B}] [\mathcal{C}] + [\mathcal{A}] [\mathcal{B}] [\mathcal{C}] .\end{aligned}$$

But the statement $\mathcal{A} \vee \mathcal{B} \vee \mathcal{C}$ is equivalent to $(\mathcal{A} \vee \mathcal{B}) \vee \mathcal{C}$. Hence, Exercise 3.12 **(a)** (applied to $\mathcal{A} \vee \mathcal{B} \vee \mathcal{C}$ and $(\mathcal{A} \vee \mathcal{B}) \vee \mathcal{C}$ instead of $\mathcal{A}$ and $\mathcal{B}$) shows that

$$\begin{aligned}[\mathcal{A} \vee \mathcal{B} \vee \mathcal{C}] &= [(\mathcal{A} \vee \mathcal{B}) \vee \mathcal{C}] \\ &= [\mathcal{A}] + [\mathcal{B}] + [\mathcal{C}] - [\mathcal{A}] [\mathcal{B}] - [\mathcal{A}] [\mathcal{C}] - [\mathcal{B}] [\mathcal{C}] + [\mathcal{A}] [\mathcal{B}] [\mathcal{C}] .\end{aligned}$$

This solves Exercise 3.12 **(e)**. □

7.23. Solution to Exercise 3.13

Proof of Theorem 3.45. Clearly, $m \geq 0$. Hence, Theorem 3.46 (applied to $k = 0$) yields

$$(-1)^m \sum_{s \in S} \dbinom{c(s)}{0} \dbinom{c(s) - 0 - 1}{m - 0} = \sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \underbrace{\dbinom{|I|}{0}}_{\substack{=1 \\ \text{(by Proposition 3.3 (a)} \\ \text{(applied to } |I| \text{ instead of } m))}} \left| \bigcap_{i \in I} A_i \right|$$
$$= \sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right|.$$

Hence,

$$\sum_{\substack{I \subseteq G; \\ |I| \leq m}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| = (-1)^m \sum_{s \in S} \underbrace{\dbinom{c(s)}{0}}_{\substack{=1 \\ \text{(by Proposition 3.3 (a)} \\ \text{(applied to } c(s) \text{ instead of } m))}} \underbrace{\dbinom{c(s) - 0 - 1}{m - 0}}_{= \dbinom{c(s) - 1}{m}}$$
$$= (-1)^m \sum_{s \in S} \dbinom{c(s) - 1}{m}.$$

This proves Theorem 3.45. $\square$

Proof of Theorem 3.44. Clearly, $|G| \in \mathbb{N}$ (since the set G is finite). Define an $m \in \mathbb{N}$ by $m = \max\{|G|, k\}$. Thus, $m = \max\{|G|, k\} \geq |G|$ and $m = \max\{|G|, k\} \geq k$. Also, every subset I of G satisfies $|I| \leq m$ [313]. Hence, we have the following equality of summation signs:

$$\sum_{\substack{I \subseteq G; \\ |I| \leq m}} = \sum_{I \subseteq G}. \tag{576}$$

For each $s \in S$, let $c(s)$ denote the number of $i \in G$ satisfying $s \in A_i$.
Recall that

$$S_k = \{s \in S \mid \text{the number of } i \in G \text{ satisfying } s \in A_i \text{ equals } k\}.$$

[313] *Proof.* Let I be a subset of G. Thus, $|I| \leq |G| \leq m$ (since $m \geq |G|$). Qed.

Hence, for each $s \in S$, we have the following chain of logical equivalences:

$$\begin{aligned}
(s \in S_k) &\iff \left(\underbrace{\text{the number of } i \in G \text{ satisfying } s \in A_i}_{\substack{=c(s) \\ \text{(since } c(s) \text{ is the number of } i \in G \text{ satisfying } s \in A_i \\ \text{(by the definition of } c(s)\text{))}}} \text{ equals } k \right) \\
&\iff (c(s) \text{ equals } k) \\
&\iff (c(s) = k) .
\end{aligned} \tag{577}$$

Observe that $m - k \in \mathbb{N}$ (since $m \geq k$).

Next, we notice the following:

Observation 1: For any $s \in S$, we have

$$\binom{c(s)}{k} \binom{c(s) - k - 1}{m - k} = (-1)^{m-k} [s \in S_k] .$$

[*Proof of Observation 1:* Let $s \in S$. Clearly, $c(s) \in \mathbb{N}$. We are in one of the following three cases:

Case 1: We have $c(s) < k$.

Case 2: We have $c(s) = k$.

Case 3: We have $c(s) > k$.

Let us first consider Case 1. In this case, we have $c(s) < k$. Thus, Proposition 3.6 (applied to $c(s)$ and k instead of m and n) yields $\binom{c(s)}{k} = 0$. Hence, $\underbrace{\binom{c(s)}{k}}_{=0} \binom{c(s) - k - 1}{m - k} = 0$.

But $c(s) \neq k$ (since $c(s) < k$). Thus, we don't have $c(s) = k$. Hence, we don't have $s \in S_k$ (by the equivalence (577)). Thus, $[s \in S_k] = 0$. Hence, $(-1)^{m-k} \underbrace{[s \in S_k]}_{=0} = 0$. Comparing this with $\binom{c(s)}{k} \binom{c(s) - k - 1}{m - k} = 0$, we obtain $\binom{c(s)}{k} \binom{c(s) - k - 1}{m - k} = (-1)^{m-k} [s \in S_k]$. Hence, Observation 1 is proven in Case 1.

Let us now consider Case 2. In this case, we have $c(s) = k$. Hence,

$$\binom{c(s)}{k} \binom{c(s) - k - 1}{m - k} = \underbrace{\binom{k}{k}}_{\substack{=1 \\ \text{(by Proposition 3.9} \\ \text{(applied to } k \text{ instead of } m\text{))}}} \underbrace{\binom{k - k - 1}{m - k}}_{= \binom{-1}{m-k}} = \binom{-1}{m - k} = (-1)^{m-k}$$

(by Corollary 3.17 (applied to $n = m - k$)).

But $c(s) = k$. Hence, $s \in S_k$ (by the equivalence (577)). Thus, $[s \in S_k] = 1$. Hence, $(-1)^{m-k} \underbrace{[s \in S_k]}_{=1} = (-1)^{m-k}$. Comparing this with $\dbinom{c(s)}{k}\dbinom{c(s)-k-1}{m-k} = (-1)^{m-k}$, we obtain $\dbinom{c(s)}{k}\dbinom{c(s)-k-1}{m-k} = (-1)^{m-k}[s \in S_k]$. Hence, Observation 1 is proven in Case 2.

Let us finally consider Case 3. In this case, we have $c(s) > k$. Thus, $c(s) \geq k+1$ (since $c(s)$ and k are integers), so that $c(s) - k \geq 1$ and thus $c(s) - k - 1 \in \mathbb{N}$. But $c(s)$ is the number of $i \in G$ satisfying $s \in A_i$. Hence,

$$c(s) = (\text{the number of } i \in G \text{ satisfying } s \in A_i) = \left|\underbrace{\{i \in G \mid s \in A_i\}}_{\subseteq G}\right| \leq |G| \leq m$$

(since $m \geq |G|$). Hence, $\underbrace{c(s)}_{\leq m} - k - 1 \leq m - k - 1 < m - k$. Therefore, Proposition 3.6 (applied to $c(s) - k - 1$ and $m - k$ instead of m and n) yields $\dbinom{c(s)-k-1}{m-k} = 0$. Hence, $\dbinom{c(s)}{k}\underbrace{\dbinom{c(s)-k-1}{m-k}}_{=0} = 0$.

But $c(s) \neq k$ (since $c(s) > k$). Thus, we don't have $c(s) = k$. Hence, we don't have $s \in S_k$ (by the equivalence (577)). Thus, $[s \in S_k] = 0$. Hence, $(-1)^{m-k} \underbrace{[s \in S_k]}_{=0} = 0$. Comparing this with $\dbinom{c(s)}{k}\dbinom{c(s)-k-1}{m-k} = 0$, we obtain $\dbinom{c(s)}{k}\dbinom{c(s)-k-1}{m-k} = (-1)^{m-k}[s \in S_k]$. Hence, Observation 1 is proven in Case 3.

We have now proven Observation 1 in each of the three Cases 1, 2 and 3. Hence, Observation 1 always holds.]

Now, Theorem 3.46 yields

$$\begin{aligned}
(-1)^m \sum_{s \in S} \binom{c(s)}{k}\binom{c(s)-k-1}{m-k} &= \underbrace{\sum_{\substack{I \subseteq G;\\ |I| \leq m}}}_{\substack{=\sum_{I \subseteq G} \\ \text{(by (576))}}} (-1)^{|I|} \binom{|I|}{k} \left|\bigcap_{i \in I} A_i\right| \\
&= \sum_{I \subseteq G} (-1)^{|I|} \binom{|I|}{k} \left|\bigcap_{i \in I} A_i\right|.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{I\subseteq G}(-1)^{|I|}\binom{|I|}{k}\left|\bigcap_{i\in I}A_i\right| \\
&=(-1)^m\sum_{s\in S}\underbrace{\binom{c(s)}{k}\binom{c(s)-k-1}{m-k}}_{\substack{=(-1)^{m-k}[s\in S_k]\\ \text{(by Observation 1)}}} \\
&=(-1)^m\sum_{s\in S}(-1)^{m-k}[s\in S_k]=\underbrace{(-1)^m(-1)^{m-k}}_{\substack{=(-1)^{m+(m-k)}=(-1)^k\\ (\text{since } m+(m-k)=2m-k\equiv -k\equiv k \bmod 2)}}\sum_{s\in S}[s\in S_k] \\
&=(-1)^k\sum_{s\in S}[s\in S_k]. \qquad (578)
\end{aligned}$$

Multiplying both sides of this equality by $(-1)^k$, we find

$$\begin{aligned}
&(-1)^k\sum_{I\subseteq G}(-1)^{|I|}\binom{|I|}{k}\left|\bigcap_{i\in I}A_i\right| \\
&=\underbrace{(-1)^k(-1)^k}_{\substack{=((-1)(-1))^k=1^k\\ (\text{since } (-1)(-1)=1)}}\sum_{s\in S}[s\in S_k]=\underbrace{1^k}_{=1}\sum_{s\in S}[s\in S_k] \\
&=\sum_{s\in S}[s\in S_k]. \qquad (579)
\end{aligned}$$

But S_k is a subset of S (by the definition of S_k). Hence, Lemma 3.49 (applied to $T=S_k$) yields

$$\begin{aligned}
|S_k|&=\sum_{s\in S}[s\in S_k]=(-1)^k\sum_{I\subseteq G}(-1)^{|I|}\binom{|I|}{k}\left|\bigcap_{i\in I}A_i\right| \qquad (\text{by (579)}) \\
&=\sum_{I\subseteq G}\underbrace{(-1)^{|I|}(-1)^k}_{\substack{=(-1)^{|I|+k}=(-1)^{|I|-k}\\ (\text{since } |I|+k\equiv |I|-k \bmod 2)}}\binom{|I|}{k}\left|\bigcap_{i\in I}A_i\right|=\sum_{I\subseteq G}(-1)^{|I|-k}\binom{|I|}{k}\left|\bigcap_{i\in I}A_i\right|.
\end{aligned}$$

This proves Theorem 3.44. □

Proof of Theorem 3.43. We have $\bigcup_{i\in G} A_i \subseteq S$ (since all the A_i are subsets of S). Hence,

$$\bigcup_{i\in G} A_i = \underbrace{\left(\bigcup_{i\in G} A_i\right)}_{\substack{=\{s \mid \text{there exists an } i\in G \text{ such that } s\in A_i\} \\ \text{(by the definition of the union } \bigcup_{i\in G} A_i)}} \cap S$$
$$= \{s \mid \text{there exists an } i \in G \text{ such that } s \in A_i\} \cap S$$
$$= \{s \in S \mid \text{there exists an } i \in G \text{ such that } s \in A_i\}. \qquad (580)$$

Define a $k \in \mathbb{N}$ by $k = 0$. Define the set S_k as in Theorem 3.44. Then,

$$S_k = \{s \in S \mid \text{the number of } i \in G \text{ satisfying } s \in A_i \text{ equals } k\}$$
$$= \left\{ s \in S \mid \underbrace{\text{the number of } i \in G \text{ satisfying } s \in A_i \text{ equals } 0}_{\iff (\text{there exists no } i\in G \text{ such that } s\in A_i)} \right\}$$
$$(\text{since } k = 0)$$
$$= \{s \in S \mid \text{there exists no } i \in G \text{ such that } s \in A_i\}$$
$$= S \setminus \underbrace{\{s \in S \mid \text{there exists an } i \in G \text{ such that } s \in A_i\}}_{\substack{=\bigcup_{i\in G} A_i \\ \text{(by (580))}}}$$
$$= S \setminus \left(\bigcup_{i\in G} A_i\right).$$

Hence, $|S_k| = \left| S \setminus \left(\bigcup_{i\in G} A_i\right) \right|$. Therefore,

$$\left| S \setminus \left(\bigcup_{i\in G} A_i\right) \right| = |S_k| = \sum_{I\subseteq G} \underbrace{(-1)^{|I|-k}}_{\substack{=(-1)^{|I|-0} \\ \text{(since } k=0)}} \underbrace{\binom{|I|}{k}}_{\substack{=\binom{|I|}{0} \\ \text{(since } k=0)}} \left| \bigcap_{i\in I} A_i \right|$$
$$(\text{by Theorem 3.44})$$
$$= \sum_{I\subseteq G} \underbrace{(-1)^{|I|-0}}_{=(-1)^{|I|}} \underbrace{\binom{|I|}{0}}_{\substack{=1 \\ \text{(by Proposition 3.3 (a)} \\ \text{(applied to } |I| \text{ instead of } m))}} \left| \bigcap_{i\in I} A_i \right|$$
$$= \sum_{I\subseteq G} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right|.$$

This proves Theorem 3.43. $\square$

Proof of Theorem 3.42. Let S denote the set $\bigcup_{i\in G} A_i$. Then, S is the union of finitely many finite sets (since the set G is finite, and since each of the sets A_i is finite), and thus itself is a finite set. Moreover, $S = \bigcup_{i\in G} A_i$; thus, A_i is a subset of S for each $i \in G$.

We define the intersection $\bigcap_{i\in\varnothing} A_i$ (which would otherwise be undefined, since $\varnothing$ is the empty set) to mean the set S. (Thus, $\bigcap_{i\in I} A_i$ is defined for any subset I of G, not just for nonempty subsets I.)

We have $\bigcap_{i\in\varnothing} A_i = S$ (since we have defined $\bigcap_{i\in\varnothing} A_i$ to be S).

Theorem 3.43 yields

$$\begin{aligned}
\left| S \setminus \left(\bigcup_{i\in G} A_i \right) \right| &= \sum_{I\subseteq G} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right| \\
&= \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right| + \underbrace{(-1)^{|\varnothing|}}_{\substack{=(-1)^0 \\ (\text{since } |\varnothing|=0)}} \left| \underbrace{\bigcap_{i\in\varnothing} A_i}_{=S} \right| \\
&\qquad \left(\begin{array}{c} \text{here, we have split off the addend for } I=\varnothing \text{ from the sum,} \\ \text{since } \varnothing \text{ is a subset of } G \end{array} \right) \\
&= \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right| + \underbrace{(-1)^0}_{=1} |S| = \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right| + |S| .
\end{aligned}$$

Comparing this with

$$\left| S \setminus \underbrace{\left(\bigcup_{i\in G} A_i \right)}_{=S} \right| = \left| \underbrace{S\setminus S}_{=\varnothing} \right| = |\varnothing| = 0,$$

we obtain

$$0 = \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right| + |S| .$$

Solving this equation for $|S|$, we find

$$|S| = -\sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|} \left| \bigcap_{i\in I} A_i \right| = \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} \underbrace{\left(-(-1)^{|I|} \right)}_{=(-1)^{|I|-1}} \left| \bigcap_{i\in I} A_i \right| = \sum_{\substack{I\subseteq G;\\ I\neq\varnothing}} (-1)^{|I|-1} \left| \bigcap_{i\in I} A_i \right| .$$

In view of $S = \bigcup_{i \in G} A_i$, this rewrites as

$$\left| \bigcup_{i \in G} A_i \right| = \sum_{\substack{I \subseteq G; \\ I \neq \varnothing}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right| .$$

This proves Theorem 3.42. □

Solution to Exercise 3.13. We have proven Theorem 3.42, Theorem 3.43, Theorem 3.44 and Theorem 3.45. Thus, Exercise 3.13 is solved. □

7.24. Solution to Exercise 3.15

Exercise 3.15 is one of the most fundamental results in combinatorics. A combinatorial proof is sketched, e.g., in [Galvin17, proof of Proposition 13.3] and in [BenQui03, proof of Identity 143]. We shall give a different proof, using induction instead.

Let us first state a basic lemma about sets:

Lemma 7.18. Let M be a positive integer. For every $i \in \{1, 2, \ldots, M\}$, let Z_i be a set. Then, the map

$$Z_1 \times Z_2 \times \cdots \times Z_M \to (Z_1 \times Z_2 \times \cdots \times Z_{M-1}) \times Z_M,$$
$$(s_1, s_2, \ldots, s_M) \mapsto ((s_1, s_2, \ldots, s_{M-1}), s_M)$$

is a bijection.

Proof of Lemma 7.18. This map is the canonical bijection $Z_1 \times Z_2 \times \cdots \times Z_M \to (Z_1 \times Z_2 \times \cdots \times Z_{M-1}) \times Z_M$. Its inverse map sends each $((s_1, s_2, \ldots, s_{M-1}), t) \in (Z_1 \times Z_2 \times \cdots \times Z_{M-1}) \times Z_M$ to $(s_1, s_2, \ldots, s_{M-1}, t) \in Z_1 \times Z_2 \times \cdots \times Z_M$. □

For the sake of convenience, let us state a particular case of Lemma 7.18:

Corollary 7.19. Let M be a positive integer. Let Z be a set. Then, the map

$$Z^M \to Z^{M-1} \times Z,$$
$$(s_1, s_2, \ldots, s_M) \mapsto ((s_1, s_2, \ldots, s_{M-1}), s_M)$$

is a bijection.

Proof of Corollary 7.19. Lemma 7.18 (applied to $Z_i = Z$) shows that the map

$$\underbrace{Z \times Z \times \cdots \times Z}_{M \text{ factors}} \to \left(\underbrace{Z \times Z \times \cdots \times Z}_{M-1 \text{ factors}} \right) \times Z,$$
$$(s_1, s_2, \ldots, s_M) \mapsto ((s_1, s_2, \ldots, s_{M-1}), s_M)$$

is a bijection. Since $\underbrace{Z \times Z \times \cdots \times Z}_{M \text{ factors}} = Z^M$ and $\underbrace{Z \times Z \times \cdots \times Z}_{M-1 \text{ factors}} = Z^{M-1}$, this can be rewritten as follows: The map

$$Z^M \to Z^{M-1} \times Z,$$
$$(s_1, s_2, \ldots, s_M) \mapsto ((s_1, s_2, \ldots, s_{M-1}), s_M)$$

is a bijection. This proves Corollary 7.19. □

Next, we state a lemma that is essentially the statement of Exercise 3.15:

Lemma 7.20. Every $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy

$$\sum_{\substack{(a_1,a_2,\ldots,a_m)\in\mathbb{N}^m;\\ a_1+a_2+\cdots+a_m=n}} 1 = \binom{n+m-1}{n}.$$

Proof of Lemma 7.20. We shall prove Lemma 7.20 by induction over m:

Induction base: Lemma 7.20 holds for $m = 0$ [314]. This completes the induction base.

[314] *Proof.* We must show that Lemma 7.20 holds for $m = 0$. In other words, we must prove that $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^0;\\ a_1+a_2+\cdots+a_0=n}} 1 = \binom{n+0-1}{n}$ for all $n \in \mathbb{N}$.

Fix $n \in \mathbb{N}$. We are in one of the following two cases:

Case 1: We have $n = 0$.

Case 2: We have $n \neq 0$.

Let us first consider Case 1. In this case, we have $n = 0$. There exists exactly one 0-tuple $(a_1, a_2, \ldots, a_0) \in \mathbb{N}^0$ (namely, the empty list $()$), and this 0-tuple satisfies $a_1 + a_2 + \cdots + a_0 = n$ (because it satisfies $a_1 + a_2 + \cdots + a_0 = (\text{empty sum}) = 0 = n$). Thus, the sum $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^0;\\ a_1+a_2+\cdots+a_0=n}} 1$ has exactly one addend (namely, the addend corresponding to $(a_1, a_2, \ldots, a_0) = ()$). Hence, this sum rewrites as follows:

$$\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^0;\\ a_1+a_2+\cdots+a_0=n}} 1 = 1.$$

Comparing this with

$$\binom{n+0-1}{n} = \binom{0+0-1}{0} \qquad (\text{since } n = 0)$$
$$= 1 \qquad (\text{by Proposition 3.3 } \textbf{(a)} \text{ (applied to } 0+0-1 \text{ instead of } m)),$$

we obtain $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^0;\\ a_1+a_2+\cdots+a_0=n}} 1 = \binom{n+0-1}{n}$. Hence, the equality $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^0;\\ a_1+a_2+\cdots+a_0=n}} 1 = \binom{n+0-1}{n}$ is proven in Case 1.

Let us now consider Case 2. In this case, we have $n \neq 0$. Hence, n is a positive integer (since $n \in \mathbb{N}$). Thus, $n - 1 \in \mathbb{N}$.

Each 0-tuple $(a_1, a_2, \ldots, a_0) \in \mathbb{N}^0$ satisfies $a_1 + a_2 + \cdots + a_0 = (\text{empty sum}) = 0 \neq n$. In other

Induction step: Fix a positive integer M. Assume that Lemma 7.20 holds for $m = M - 1$. We now must show that Lemma 7.20 holds for $m = M$.

We have assumed that Lemma 7.20 holds for $m = M - 1$. In other words, every $n \in \mathbb{N}$ satisfies

$$\sum_{\substack{(a_1,a_2,\ldots,a_{M-1})\in\mathbb{N}^{M-1};\\ a_1+a_2+\cdots+a_{M-1}=n}} 1 = \binom{n+(M-1)-1}{n}. \qquad (581)$$

We must show that Lemma 7.20 holds for $m = M$. In other words, we must prove that every $n \in \mathbb{N}$ satisfies

$$\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^{M};\\ a_1+a_2+\cdots+a_M=n}} 1 = \binom{n+M-1}{n}. \qquad (582)$$

Let us now prove this.

Let $n \in \mathbb{N}$. For each $(a_1, a_2, \ldots, a_M) \in \mathbb{N}^M$ satisfying $a_1 + a_2 + \cdots + a_M = n$, we have $n - a_M \in \{0, 1, \ldots, n\}$ [315]. Hence, we have the following equality of

---

words, no $(a_1, a_2, \ldots, a_0) \in \mathbb{N}^0$ satisfies $a_1 + a_2 + \cdots + a_0 = n$. Hence, the sum $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^{0};\\ a_1+a_2+\cdots+a_0=n}} 1$ is an empty sum. Therefore,

$$\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^{0};\\ a_1+a_2+\cdots+a_0=n}} 1 = (\text{empty sum}) = 0.$$

But $n - 1 < n$ and $n - 1 \in \mathbb{N}$. Thus, Proposition 3.6 (applied to $n - 1$ instead of m) yields $\binom{n-1}{n} = 0$. Hence, $\binom{n+0-1}{n} = \binom{n-1}{n} = 0$. Thus, $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^{0};\\ a_1+a_2+\cdots+a_0=n}} 1 = 0 = \binom{n+0-1}{n}$.

Hence, the equality $\sum_{\substack{(a_1,a_2,\ldots,a_0)\in\mathbb{N}^{0};\\ a_1+a_2+\cdots+a_0=n}} 1 = \binom{n+0-1}{n}$ is proven in Case 2.

We have now proven this equality in each of the two Cases 1 and 2. Thus, this equality always holds. In other words, Lemma 7.20 holds for $m = 0$.

[315] *Proof.* Let $(a_1, a_2, \ldots, a_M) \in \mathbb{N}^M$ be such that $a_1 + a_2 + \cdots + a_M = n$. We must show that $n - a_M \in \{0, 1, \ldots, n\}$.

We have $(a_1, a_2, \ldots, a_M) \in \mathbb{N}^M$. Thus, $a_1, a_2, \ldots, a_M$ are elements of $\mathbb{N}$. Hence, in particular, $a_1, a_2, \ldots, a_{M-1}$ are elements of $\mathbb{N}$. Thus, $a_1 + a_2 + \cdots + a_{M-1} \in \mathbb{N}$, so that $a_1 + a_2 + \cdots + a_{M-1} \geq 0$.

From $a_1 + a_2 + \cdots + a_M = n$, we obtain

$$n = a_1 + a_2 + \cdots + a_M = \underbrace{(a_1 + a_2 + \cdots + a_{M-1})}_{\geq 0} + a_M \qquad (\text{since } M \text{ is a positive integer})$$
$$\geq a_M,$$

so that $n - a_M \geq 0$. Also, we know that $a_1, a_2, \ldots, a_M$ are elements of $\mathbb{N}$; thus, $a_M \in \mathbb{N}$. Hence, $a_M \geq 0$.

Clearly, n and a_M are integers (since $n \in \mathbb{N} \subseteq \mathbb{Z}$ and $a_M \in \mathbb{N} \subseteq \mathbb{Z}$). Thus, $n - a_M$ is an integer. Combining this with $n - a_M \geq 0$, we conclude that $n - a_M \in \mathbb{N}$. Combining this with summation signs:

$$\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n}} = \underbrace{\sum_{r\in\{0,1,\ldots,n\}}}_{=\sum_{r=0}^{n}} \sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}} = \sum_{r=0}^{n} \sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}} . \tag{583}$$

Now, we are going to show that

$$\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}} 1 = \binom{r+M-2}{r} \tag{584}$$

for each $r \in \{0,1,\ldots,n\}$.

[*Proof of (584):* Let $r \in \{0,1,\ldots,n\}$.

For any $(a_1,a_2,\ldots,a_M) \in \mathbb{N}^M$, the condition $(a_1+a_2+\cdots+a_M = n$ and $n - a_M = r)$ is equivalent to the condition $(a_1+a_2+\cdots+a_{M-1} = r$ and $a_M = n-r)$. [316]

Thus, we have the following equality of summation signs:

$$\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}} = \sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_{M-1}=r;\\ a_M=n-r}} . \tag{585}$$

We have $r \leq n$ (since $r \in \{0,1,\ldots,n\}$), thus $n \geq r$ and therefore $n-r \geq 0$. Hence, $n-r \in \mathbb{N}$. Thus, the sum $\sum_{\substack{q\in\mathbb{N};\\ q=n-r}} 1$ has exactly one addend (namely, the addend for $q = n-r$). Hence, it simplifies as follows:

$$\sum_{\substack{q\in\mathbb{N};\\ q=n-r}} 1 = 1. \tag{586}$$

$n - \underbrace{a_M}_{\geq 0} \leq n$, we obtain $n - a_M \in \{0,1,\ldots,n\}$. Qed.

[316]This can be checked easily. For example, in order to prove the implication

$$(a_1+a_2+\cdots+a_M = n \text{ and } n-a_M = r) \implies (a_1+a_2+\cdots+a_{M-1} = r \text{ and } a_M = n-r),$$

it suffices to assume that $(a_1+a_2+\cdots+a_M = n$ and $n-a_M = r)$ holds, and then to conclude that

$$a_1+a_2+\cdots+a_{M-1} = \underbrace{(a_1+a_2+\cdots+a_M)}_{=n} - a_M = n - a_M = r \qquad \text{and}$$

$$a_M = n - r \qquad (\text{since } n - a_M = r).$$

The converse implication is proven similarly.

Corollary 7.19 (applied to $Z = \mathbb{N}$) shows that the map

$$\mathbb{N}^M \to \mathbb{N}^{M-1} \times \mathbb{N},$$
$$(s_1, s_2, \ldots, s_M) \mapsto ((s_1, s_2, \ldots, s_{M-1}), s_M)$$

is a bijection. Hence, we can substitute $((s_1, s_2, \ldots, s_{M-1}), s_M)$ for $((a_1, a_2, \ldots, a_{M-1}), q)$ in the sum $\sum\limits_{\substack{((a_1,a_2,\ldots,a_{M-1}),q)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ a_1+a_2+\cdots+a_{M-1}=r;\\ q=n-r}} 1$. Thus, we obtain

$$\sum_{\substack{((a_1,a_2,\ldots,a_{M-1}),q)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ a_1+a_2+\cdots+a_{M-1}=r;\\ q=n-r}} 1 = \sum_{\substack{(s_1,s_2,\ldots,s_M)\in\mathbb{N}^M;\\ s_1+s_2+\cdots+s_{M-1}=r;\\ s_M=n-r}} 1 = \sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_{M-1}=r;\\ a_M=n-r}} 1$$

(here, we have renamed the summation index $(s_1, s_2, \ldots, s_M)$ as $(a_1, a_2, \ldots, a_M)$). Hence,

$$\begin{aligned}
\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_{M-1}=r;\\ a_M=n-r}} 1 &= \underbrace{\sum_{\substack{((a_1,a_2,\ldots,a_{M-1}),q)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ a_1+a_2+\cdots+a_{M-1}=r;\\ q=n-r}}}_{=\sum\limits_{\substack{(a_1,a_2,\ldots,a_{M-1})\in\mathbb{N}^{M-1};\\ a_1+a_2+\cdots+a_{M-1}=r}} \sum\limits_{\substack{q\in\mathbb{N};\\ q=n-r}}} 1 = \sum_{\substack{(a_1,a_2,\ldots,a_{M-1})\in\mathbb{N}^{M-1};\\ a_1+a_2+\cdots+a_{M-1}=r}} \underbrace{\sum_{\substack{q\in\mathbb{N};\\ q=n-r}} 1}_{\substack{=1\\ \text{(by (586))}}} \\
&= \sum_{\substack{(a_1,a_2,\ldots,a_{M-1})\in\mathbb{N}^{M-1};\\ a_1+a_2+\cdots+a_{M-1}=r}} 1 = \binom{r+(M-1)-1}{r} \\
&\qquad \text{(by (581) (applied to } r \text{ instead of } n\text{))} \\
&= \binom{r+M-2}{r}
\end{aligned}$$

(since $(M-1)-1 = M-2$). Now,

$$\underbrace{\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}}}_{\substack{=\sum\limits_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_{M-1}=r;\\ a_M=n-r}}\\ \text{(by (585))}}} 1 = \sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_{M-1}=r;\\ a_M=n-r}} 1 = \binom{r+M-2}{r}.$$

Thus, (584) is proven.]

Now,

$$\underbrace{\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n}} 1}_{\substack{=\sum\limits_{r=0}^{n} \sum\limits_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}} \\ \text{(by (583))}}} = \sum_{r=0}^{n} \underbrace{\sum_{\substack{(a_1,a_2,\ldots,a_M)\in\mathbb{N}^M;\\ a_1+a_2+\cdots+a_M=n;\\ n-a_M=r}} 1}_{\substack{=\dbinom{r+M-2}{r} \\ \text{(by (584))}}} = \sum_{r=0}^{n} \dbinom{r+M-2}{r}$$

$$= \dbinom{n+(M-2)+1}{n} \qquad \text{(by Lemma 7.9 (applied to } q = M-2\text{))}$$

$$= \dbinom{n+M-1}{n} \qquad \left(\text{since } (M-2)+1 = M-1\right).$$

In other words, (582) holds.

Now, forget that we fixed n. We thus have proven that every $n \in \mathbb{N}$ satisfies (582). In other words, Lemma 7.20 holds for $m = M$. This completes the induction step. The induction proof of Lemma 7.20 is thus complete. □

Solution to Exercise 3.15. Lemma 7.20 (applied to $m = k$) yields

$$\sum_{\substack{(a_1,a_2,\ldots,a_k)\in\mathbb{N}^k;\\ a_1+a_2+\cdots+a_k=n}} 1 = \dbinom{n+k-1}{n}.$$

Comparing this with

$$\sum_{\substack{(a_1,a_2,\ldots,a_k)\in\mathbb{N}^k;\\ a_1+a_2+\cdots+a_k=n}} 1$$
$$= \left|\left\{(a_1,a_2,\ldots,a_k) \in \mathbb{N}^k \mid a_1+a_2+\cdots+a_k = n\right\}\right| \cdot 1$$
$$= \left|\left\{(a_1,a_2,\ldots,a_k) \in \mathbb{N}^k \mid a_1+a_2+\cdots+a_k = n\right\}\right|$$
$$= \left(\text{the number of all } k\text{-tuples } (a_1,a_2,\ldots,a_k) \in \mathbb{N}^k \text{ satisfying } a_1+a_2+\cdots+a_k = n\right),$$

we obtain

$$\left(\text{the number of all } k\text{-tuples } (a_1,a_2,\ldots,a_k) \in \mathbb{N}^k \text{ satisfying } a_1+a_2+\cdots+a_k = n\right)$$
$$= \dbinom{n+k-1}{n}.$$

In other words, the number of all k-tuples $(a_1,a_2,\ldots,a_k) \in \mathbb{N}^k$ satisfying $a_1+a_2+\cdots+a_k = n$ equals $\dbinom{n+k-1}{n}$. This solves Exercise 3.15. □

7.25. Solution to Exercise 3.16

Exercise 3.16 is an identity that tends to creep up in various seemingly unrelated situations in mathematics. I have first encountered it in [Schmit04, proof of Theorem 9.5] (where it appears with an incorrect power of -1 on the right hand side). It also has recently appeared on math.stackexchange ([dilemi17], with a, $n-a$ and $k-a$ renamed as p, n and k), where it has been proven in three different ways: once using the beta function, once using residues, and once (by myself in the comments) using finite differences. Let me here give a different, elementary proof.

We begin with the following identities:

Proposition 7.21. Let $i \in \mathbb{Z}$, $n \in \mathbb{Z}$ and $j \in \mathbb{N}$. Then:

(a) We have

$$\sum_{k=0}^{j} (-1)^k \binom{n}{j-k} \binom{k+i-1}{k} = \binom{n-i}{j}.$$

(b) If i is positive, then

$$\sum_{k=0}^{j} \frac{(-1)^k}{k+i} \binom{n}{j-k} \binom{k+i}{i} = \frac{1}{i} \binom{n-i}{j}.$$

Proof of Proposition 7.21. **(a)** Proposition 3.32 **(a)** (applied to $-i$, n and j instead of x, y and n) yields

$$\binom{(-i)+n}{j} = \sum_{k=0}^{j} \underbrace{\binom{-i}{k}}_{\substack{=(-1)^k \binom{k-(-i)-1}{k} \\ \text{(by Proposition 3.16,} \\ \text{applied to } -i \text{ and } k \text{ instead of } m \text{ and } n)}} \binom{n}{j-k}$$

$$= \sum_{k=0}^{j} (-1)^k \underbrace{\binom{k-(-i)-1}{k}}_{\substack{=\binom{k+i-1}{k} \\ \text{(since } k-(-i)-1=k+i-1)}} \binom{n}{j-k}$$

$$= \sum_{k=0}^{j} (-1)^k \binom{k+i-1}{k} \binom{n}{j-k} = \sum_{k=0}^{j} (-1)^k \binom{n}{j-k} \binom{k+i-1}{k}.$$

Thus,

$$\sum_{k=0}^{j} (-1)^k \binom{n}{j-k} \binom{k+i-1}{k} = \binom{(-i)+n}{j} = \binom{n-i}{j}.$$

This proves Proposition 7.21 **(a)**.

(b) Assume that i is positive. Let $k \in \mathbb{N}$. Then, $i-1 \in \mathbb{N}$ (since i is a positive integer). Thus, $i-1 \geq 0$. Also, $\underbrace{k}_{\geq 0} + i - 1 \geq i-1$. Hence, Proposition 3.8 (applied to $k+i-1$ and $i-1$ instead of m and n) yields

$$\binom{k+i-1}{i-1} = \binom{k+i-1}{(k+i-1)-(i-1)} = \binom{k+i-1}{k}$$

(since $(k+i-1)-(i-1) = k$).

Furthermore, $i \in \{1,2,3,\ldots\}$ (since i is a positive integer). Thus, Proposition 3.22 (applied to $k+i$ and i instead of m and n) yields $\binom{k+i}{i} = \frac{k+i}{i}\binom{k+i-1}{i-1}$. Multiplying both sides of this equality by i, we find

$$i\binom{k+i}{i} = i \cdot \frac{k+i}{i}\binom{k+i-1}{i-1} = (k+i)\underbrace{\binom{k+i-1}{i-1}}_{=\binom{k+i-1}{k}} = (k+i)\binom{k+i-1}{k}.$$

Hence,

$$\binom{k+i}{i} = \frac{1}{i}(k+i)\binom{k+i-1}{k}.$$

Thus,

$$\begin{aligned}
&\frac{(-1)^k}{k+i}\binom{n}{j-k}\underbrace{\binom{k+i}{i}}_{=\frac{1}{i}(k+i)\binom{k+i-1}{k}} \\
&= \frac{(-1)^k}{k+i}\binom{n}{j-k} \cdot \frac{1}{i}(k+i)\binom{k+i-1}{k} \\
&= \frac{1}{i}(-1)^k\binom{n}{j-k}\binom{k+i-1}{k}.
\end{aligned} \tag{587}$$

Now, forget that we fixed k. We thus have proven (587) for each $k \in \mathbb{N}$. Now,

$$\sum_{k=0}^{j} \underbrace{\frac{(-1)^k}{k+i}\binom{n}{j-k}\binom{k+i}{i}}_{\substack{=\frac{1}{i}(-1)^k\binom{n}{j-k}\binom{k+i-1}{k} \\ \text{(by (587))}}}$$
$$= \sum_{k=0}^{j} \frac{1}{i}(-1)^k \binom{n}{j-k}\binom{k+i-1}{k}$$
$$= \frac{1}{i} \underbrace{\sum_{k=0}^{j} (-1)^k \binom{n}{j-k}\binom{k+i-1}{k}}_{\substack{=\binom{n-i}{j} \\ \text{(by Proposition 7.21 (a))}}} = \frac{1}{i}\binom{n-i}{j}.$$

This proves Proposition 7.21 **(b)**. □

Let us now solve the actual exercise:

Solution to Exercise 3.16. From $n \geq a$, we obtain $n - a \in \mathbb{N}$. Also, $n \geq a \geq 1 \geq 0$, and therefore Proposition 3.4 (applied to n and a instead of m and n) yields $\binom{n}{a} = \frac{n!}{a!(n-a)!} \neq 0$ (since $n! \neq 0$).

Any $k \in \{a, a+1, \ldots, n\}$ satisfies

$$\binom{n}{a}\binom{n-a}{k-a} = \binom{n}{n-k}\binom{k}{a} \tag{588}$$

[317].

[317] *Proof of (588):* Let $k \in \{a, a+1, \ldots, n\}$. Thus, $a \leq k \leq n$, so that $k \geq a \geq 1 \geq 0$, so that $k \in \mathbb{N}$. Hence, Proposition 3.8 (applied to n and k instead of m and n) yields $\binom{n}{k} = \binom{n}{n-k}$ (since $n \geq k$).

But Proposition 3.23 (applied to n and k instead of m and i) shows that

$$\binom{n}{k}\binom{k}{a} = \binom{n}{a}\binom{n-a}{k-a}.$$

Hence,

$$\binom{n}{a}\binom{n-a}{k-a} = \underbrace{\binom{n}{k}}_{=\binom{n}{n-k}}\binom{k}{a} = \binom{n}{n-k}\binom{k}{a}.$$

This proves (588).

We have

$$\begin{aligned}
\binom{n}{a}\sum_{k=a}^{n}\frac{(-1)^k}{k}\binom{n-a}{k-a} &= \sum_{k=a}^{n}\frac{(-1)^k}{k}\underbrace{\binom{n}{a}\binom{n-a}{k-a}}_{\substack{=\binom{n}{n-k}\binom{k}{a} \\ \text{(by (588))}}} = \sum_{k=a}^{n}\frac{(-1)^k}{k}\binom{n}{n-k}\binom{k}{a} \\
&= \sum_{k=0}^{n-a}\underbrace{\frac{(-1)^{k+a}}{k+a}}_{\substack{=\frac{(-1)^k(-1)^a}{k+a} \\ \text{(since } (-1)^{k+a}=(-1)^k(-1)^a\text{)}}}\underbrace{\binom{n}{n-(k+a)}}_{\substack{=\binom{n}{(n-a)-k} \\ \text{(since } n-(k+a)=(n-a)-k\text{)}}}\binom{k+a}{a} \\
&\qquad (\text{here, we have substituted } k+a \text{ for } k \text{ in the sum}) \\
&= \sum_{k=0}^{n-a}\frac{(-1)^k(-1)^a}{k+a}\binom{n}{(n-a)-k}\binom{k+a}{a} \\
&= (-1)^a\underbrace{\sum_{k=0}^{n-a}\frac{(-1)^k}{k+a}\binom{n}{(n-a)-k}\binom{k+a}{a}}_{\substack{=\frac{1}{a}\binom{n-a}{n-a} \\ \text{(by Proposition 7.21 (b),} \\ \text{applied to } j=n-a \text{ and } i=a\text{)}}} \\
&= (-1)^a\frac{1}{a}\underbrace{\binom{n-a}{n-a}}_{\substack{=1 \\ \text{(by Proposition 3.9} \\ \text{(applied to } m=n-a\text{))}}} = (-1)^a\frac{1}{a}.
\end{aligned}$$

We can divide both sides of this equality by $\binom{n}{a}$ (since $\binom{n}{a} \neq 0$). Thus, we find

$$\sum_{k=a}^{n}\frac{(-1)^k}{k}\binom{n-a}{k-a} = (-1)^a\frac{1}{a}/\binom{n}{a} = \frac{(-1)^a}{a\binom{n}{a}}.$$

This solves Exercise 3.16. □

7.26. Solution to Exercise 3.18

We shall now prepare for the solution of Exercise 3.18.

Let us first fix some notations.

Definition 7.22. Let $N \in \mathbb{N}$. We shall consider N to be fixed for the whole Section 7.26.

Throughout Section 7.26, we shall use the word "list" for an $(N+1)$-tuple of rational numbers. In other words, a "list" will mean an element of $\mathbb{Q}^{N+1}$. Thus, for example, when we say "the list $(1,1,\ldots,1)$", we mean the list $\left(\underbrace{1,1,\ldots,1}_{N+1 \text{ entries}}\right)$ (because any list has to be an $(N+1)$-tuple).

Definition 7.23. The *binomial transform* of a list $(f_0, f_1, \ldots, f_N) \in \mathbb{Q}^{N+1}$ is defined to be the list $(g_0, g_1, \ldots, g_N)$ defined by

$$\left(g_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} f_i \qquad \text{for every } n \in \{0,1,\ldots,N\}\right).$$

Clearly, Definition 7.23 generalizes the definition of the binomial transform we gave in Exercise 3.18 (because any finite sequence $(f_0, f_1, \ldots, f_N) \in \mathbb{Z}^{N+1}$ of integers is clearly a list in $\mathbb{Q}^{N+1}$).

We shall use the Iverson bracket notation introduced in Definition 3.48.

The following fact is easy:

Proposition 7.24. Let $m \in \mathbb{N}$. Then,

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} = [m=0].$$

Proof of Proposition 7.24. Proposition 3.39 **(c)** (applied to $n = m$) yields

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} = \begin{cases} 1, & \text{if } m = 0; \\ 0, & \text{if } m \neq 0 \end{cases} = \begin{cases} 1, & \text{if } m = 0 \text{ is true;} \\ 0, & \text{if } m = 0 \text{ is false} \end{cases}.$$

Comparing this with

$$[m=0] = \begin{cases} 1, & \text{if } m = 0 \text{ is true;} \\ 0, & \text{if } m = 0 \text{ is false} \end{cases} \qquad (\text{by the definition of } [m=0]),$$

we obtain $\sum\limits_{k=0}^{m} (-1)^k \binom{m}{k} = [m=0]$. This proves Proposition 7.24. □

Corollary 7.25. Let $n \in \mathbb{N}$. Let $i \in \{0,1,\ldots,n\}$. Then,

$$\sum_{j=i}^{n} (-1)^{j+i} \binom{n}{j} \binom{j}{i} = [i=n].$$

Proof of Corollary 7.25. We have $i \in \{0, 1, \ldots, n\}$. Hence, $i \leq n$. Also, $i \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$. From $i \leq n$, we obtain $n - i \geq 0$. Thus, $n - i \in \mathbb{N}$. Therefore, Proposition 7.24 (applied to $m = n - i$) yields

$$\sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} = [n - i = 0] . \tag{589}$$

Let $j \in \{i, i+1, \ldots, n\}$. Then, $j \geq i$, so that $j \geq i \geq 0$ and thus $j \in \mathbb{N}$. Hence, Proposition 3.23 (applied to n, j and i instead of m, i and a) yields

$$\binom{n}{j} \binom{j}{i} = \binom{n}{i} \binom{n-i}{j-i}. \tag{590}$$

Also, $j + i \equiv j - i \mod 2$ (since $(j+i) - (j-i) = 2i$ is even). Thus, $(-1)^{j+i} = (-1)^{j-i}$. Multiplying this equality with the equality (590), we obtain

$$(-1)^{j+i} \binom{n}{j} \binom{j}{i} = (-1)^{j-i} \binom{n}{i} \binom{n-i}{j-i}. \tag{591}$$

Now, forget that we fixed j. We thus have proven (591) for each $j \in \{i, i+1, \ldots, n\}$. Hence,

$$\begin{aligned} \sum_{j=i}^{n} \underbrace{(-1)^{j+i} \binom{n}{j} \binom{j}{i}}_{\substack{=(-1)^{j-i} \binom{n}{i} \binom{n-i}{j-i} \\ \text{(by (591))}}} &= \sum_{j=i}^{n} (-1)^{j-i} \binom{n}{i} \binom{n-i}{j-i} = \sum_{k=0}^{n-i} (-1)^k \binom{n}{i} \binom{n-i}{k} \\ &\qquad \left(\text{here, we have substituted } k \text{ for } j-i \text{ in the sum}\right) \\ &= \binom{n}{i} \underbrace{\sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k}}_{\substack{=[n-i=0] \\ \text{(by (589))}}} = \binom{n}{i} [n - i = 0] . \end{aligned} \tag{592}$$

But it is easy to see that

$$\binom{n}{i} [n - i = 0] = [i = n] \tag{593}$$

[318]. Thus, (592) becomes

$$\sum_{j=i}^{n} (-1)^{j+i} \binom{n}{j} \binom{j}{i} = \binom{n}{i} [n - i = 0] = [i = n] .$$

This proves Corollary 7.25. □

[318] *Proof of (593):* We are in one of the following two cases:

Case 1: We have $i \neq n$.

Proposition 7.26. Let $(a_0, a_1, \ldots, a_N)$ and $(b_0, b_1, \ldots, b_N)$ be two $(N+1)$-tuples of rational numbers. Assume that

$$b_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} a_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}.$$

Then,

$$a_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} b_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}.$$

Proof of Proposition 7.26. We have assumed that

$$b_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} a_i \tag{594}$$

for each $n \in \{0, 1, \ldots, N\}$.

Case 2: We have $i = n$.

Let us consider Case 1 first. In this case, we have $i \neq n$. In other words, $n \neq i$. Hence, $n - i \neq 0$. Thus, the statement $n - i = 0$ is false. Hence, $[n - i = 0] = 0$, so that $\binom{n}{i} \underbrace{[n - i = 0]}_{=0} = 0$. Comparing this with $[i = n] = 0$ (since we don't have $i = n$ (since $i \neq n$)), we obtain $\binom{n}{i} [n - i = 0] = [i = n]$. Hence, (593) is proven in Case 1.

Now, let us consider Case 2. In this case, we have $i = n$. In other words, $n = i$. Thus, $n - i = 0$. Thus, $[n - i = 0] = 1$. Also, from $n = i$, we obtain $\binom{n}{i} = \binom{i}{i} = 1$ (by Proposition 3.9 (applied to $m = i$)). Hence, $\underbrace{\binom{n}{i}}_{=1} \underbrace{[n - i = 0]}_{=1} = 1$. Comparing this with $[i = n] = 1$ (since $i = n$), we obtain $\binom{n}{i} [n - i = 0] = [i = n]$. Hence, (593) is proven in Case 2.

We thus have proven (593) in each of the two Cases 1 and 2. Thus, (593) always holds.

Now, let $n \in \{0, 1, \ldots, N\}$. Then,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} b_i$$
$$= \sum_{j=0}^{n} (-1)^j \binom{n}{j} \underbrace{b_j}_{\substack{=\sum_{i=0}^{j} (-1)^i \binom{j}{i} a_i \\ \text{(by (594) (applied to } j \\ \text{instead of } n))}} \qquad \left(\begin{array}{c}\text{here, we have renamed the} \\ \text{summation index } i \text{ as } j\end{array}\right)$$
$$= \sum_{j=0}^{n} (-1)^j \binom{n}{j} \sum_{i=0}^{j} (-1)^i \binom{j}{i} a_i$$
$$= \underbrace{\sum_{j=0}^{n}}_{= \sum_{j \in \{0,1,\ldots,n\}}} \underbrace{\sum_{i=0}^{j}}_{\substack{= \sum_{i \in \{0,1,\ldots,j\}} = \sum_{i \in \{0,1,\ldots,n\}; \; i \leq j} \\ \text{(since the elements of } \{0,1,\ldots,j\} \\ \text{are precisely the elements } i \in \{0,1,\ldots,n\} \\ \text{satisfying } i \leq j \text{ (because } j \leq n))}} \underbrace{(-1)^j \binom{n}{j} (-1)^i}_{=(-1)^i \binom{n}{j}} \binom{j}{i} a_i$$
$$= \underbrace{\sum_{j \in \{0,1,\ldots,n\}} \sum_{\substack{i \in \{0,1,\ldots,n\}; \\ i \leq j}}}_{= \sum_{i \in \{0,1,\ldots,n\}} \sum_{j \in \{0,1,\ldots,n\}; \; i \leq j}} \underbrace{(-1)^j (-1)^i}_{=(-1)^{j+i}} \binom{n}{j} \binom{j}{i} a_i$$
$$= \sum_{i \in \{0,1,\ldots,n\}} \underbrace{\sum_{\substack{j \in \{0,1,\ldots,n\}; \\ i \leq j}}}_{= \sum_{j=i}^{n}} (-1)^{j+i} \binom{n}{j} \binom{j}{i} a_i$$
$$= \sum_{i \in \{0,1,\ldots,n\}} \sum_{j=i}^{n} (-1)^{j+i} \binom{n}{j} \binom{j}{i} a_i = \sum_{i \in \{0,1,\ldots,n\}} \underbrace{\left(\sum_{j=i}^{n} (-1)^{j+i} \binom{n}{j} \binom{j}{i} \right)}_{\substack{=[i=n] \\ \text{(by Corollary 7.25)}}} a_i$$
$$= \sum_{i \in \{0,1,\ldots,n\}} [i = n] a_i = \underbrace{[n = n]}_{\substack{=1 \\ \text{(since } n=n)}} a_n + \sum_{\substack{i \in \{0,1,\ldots,n\}; \\ i \neq n}} \underbrace{[i = n]}_{\substack{=0 \\ \text{(since } i \neq n)}} a_i$$

(here, we have split off the addend for $i = n$ from the sum)

$$= a_n + \underbrace{\sum_{\substack{i \in \{0,1,\ldots,n\}; \\ i \neq n}} 0 a_i}_{=0} = a_n.$$

In other words, $a_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} b_i$. This proves Proposition 7.26. □

From this, it is easy to derive the following statement, which generalizes Exercise 3.18 **(a)**:

Corollary 7.27. Let $(f_0, f_1, \ldots, f_N) \in \mathbb{Q}^{N+1}$ be a list. Let $(g_0, g_1, \ldots, g_N)$ be the binomial transform of $(f_0, f_1, \ldots, f_N)$. Then, $(f_0, f_1, \ldots, f_N)$ is the binomial transform of $(g_0, g_1, \ldots, g_N)$.

Proof of Corollary 7.27. We know that $(g_0, g_1, \ldots, g_N)$ is the binomial transform of $(f_0, f_1, \ldots, f_N)$. In other words, we have

$$g_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} f_i \qquad \text{for every } n \in \{0, 1, \ldots, N\}$$

(by the definition of the binomial transform of a list). Hence, Proposition 7.26 (applied to $a_i = f_i$ and $b_i = g_i$) yields that

$$f_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} g_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}. \tag{595}$$

Let $(h_0, h_1, \ldots, h_N)$ be the binomial transform of $(g_0, g_1, \ldots, g_N)$. Thus,

$$h_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} g_i \qquad \text{for every } n \in \{0, 1, \ldots, N\} \tag{596}$$

(by the definition of the binomial transform of a list). Hence, each $n \in \{0, 1, \ldots, N\}$ satisfies

$$h_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} g_i = f_n \qquad \text{(by (595))}.$$

In other words, $(h_0, h_1, \ldots, h_N) = (f_0, f_1, \ldots, f_N)$.

Recall that $(h_0, h_1, \ldots, h_N)$ is the binomial transform of $(g_0, g_1, \ldots, g_N)$. In other words, $(f_0, f_1, \ldots, f_N)$ is the binomial transform of $(g_0, g_1, \ldots, g_N)$ (since $(h_0, h_1, \ldots, h_N) = (f_0, f_1, \ldots, f_N)$). This proves Corollary 7.27. □

Next, let us state a simple consequence of the binomial formula:

Corollary 7.28. Let $n \in \mathbb{N}$ and $q \in \mathbb{Q}$. Then,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} q^i = (1-q)^n.$$

Proof of Corollary 7.28. Proposition 3.21 (applied to $x = -q$ and $y = 1$) yields

$$\begin{aligned}((-q)+1)^n &= \sum_{k=0}^{n} \binom{n}{k} (-q)^k \underbrace{1^{n-k}}_{=1} = \sum_{k=0}^{n} \binom{n}{k} \underbrace{(-q)^k}_{=(-1)^k q^k} \\ &= \sum_{k=0}^{n} \underbrace{\binom{n}{k} (-1)^k}_{=(-1)^k \binom{n}{k}} q^k = \sum_{k=0}^{n} (-1)^k \binom{n}{k} q^k = \sum_{i=0}^{n} (-1)^i \binom{n}{i} q^i\end{aligned}$$

(here, we have renamed the summation index k as i). Hence,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} q^i = \Bigl(\underbrace{(-q)+1}_{=1-q} \Bigr)^n = (1-q)^n .$$

This proves Corollary 7.28. □

Let us next derive an easy corollary from Corollary 7.25:

Corollary 7.29. Let $n \in \mathbb{N}$. Let $i \in \mathbb{N}$. Then,

$$\sum_{j=0}^{n} (-1)^j \binom{n}{j} \binom{j}{i} = (-1)^i [n = i] .$$

Proof of Corollary 7.29. Each $j \in \{0, 1, \ldots, i-1\}$ satisfies

$$\binom{j}{i} = 0 \qquad (597)$$

[319].

We are in one of the following two cases:

Case 1: We have $i \leq n$.

Case 2: We have $i > n$.

Let us first consider Case 1. In this case, we have $i \leq n$. Thus, $i \in \{0, 1, \ldots, n\}$.

[319] *Proof of (597):* Let $j \in \{0, 1, \ldots, i-1\}$. Thus, $j \leq i - 1 < i$ and $j \in \{0, 1, \ldots, i-1\} \subseteq \mathbb{N}$. Hence, Proposition 3.6 (applied to j and i instead of m and n) shows that $\binom{j}{i} = 0$. This proves (597).

Now,

$$\sum_{j=0}^{n}(-1)^j\binom{n}{j}\binom{j}{i}$$
$$=\sum_{j=0}^{i-1}(-1)^j\binom{n}{j}\underbrace{\binom{j}{i}}_{\substack{=0\\ \text{(by (597))}}}+\sum_{j=i}^{n}\underbrace{(-1)^j}_{\substack{=(-1)^{i+(j+i)}\\ \text{(since } j\equiv 2i+j=i+(j+i)\bmod 2)}}\binom{n}{j}\binom{j}{i}$$
$$\text{(here, we have split the sum at } j=i\text{, since } 0\leq i\leq n)$$
$$=\underbrace{\sum_{j=0}^{i-1}(-1)^j\binom{n}{j}0}_{=0}+\sum_{j=i}^{n}(-1)^{i+(j+i)}\binom{n}{j}\binom{j}{i}=\sum_{j=i}^{n}\underbrace{(-1)^{i+(j+i)}}_{=(-1)^i(-1)^{j+i}}\binom{n}{j}\binom{j}{i}$$
$$=\sum_{j=i}^{n}(-1)^i(-1)^{j+i}\binom{n}{j}\binom{j}{i}=(-1)^i\underbrace{\sum_{j=i}^{n}(-1)^{j+i}\binom{n}{j}\binom{j}{i}}_{\substack{=[i=n]\\ \text{(by Corollary 7.25)}}}$$
$$=(-1)^i\left[\underbrace{i=n}_{\Longleftrightarrow (n=i)}\right]=(-1)^i[n=i].$$

Hence, Corollary 7.29 is proven in Case 1.

Let us now consider Case 2. In this case, we have $i>n$. Thus, $n<i$, so that $n\in\{0,1,\ldots,i-1\}$ (since $n\in\mathbb{N}$). But we don't have $n=i$ (since we have $n<i$); thus, we have $[n=i]=0$. Hence, $(-1)^i\underbrace{[n=i]}_{=0}=0$. But

$$\sum_{j=0}^{n}(-1)^j\binom{n}{j}\underbrace{\binom{j}{i}}_{\substack{=0\\ \text{(by (597)}\\ \text{(since } j\in\{0,1,\ldots,i-1\}\\ \text{(because } j\leq n<i \text{ and } j\in\mathbb{N})))}}=\sum_{j=0}^{n}(-1)^j\binom{n}{j}0=0=(-1)^i[n=i].$$

Hence, Corollary 7.29 is proven in Case 2.

We have now proven Corollary 7.29 in both Cases 1 and 2. Hence, Corollary 7.29 always holds. □

Another simple lemma shall be of use to us:

Lemma 7.30. Let $n\in\mathbb{N}$.

(a) We have $\dfrac{1}{2}\left(1+(-1)^n\right)=\begin{cases}1, & \text{if } n \text{ is even;}\\ 0, & \text{if } n \text{ is odd}\end{cases}$.

(b) We have $\frac{1}{2}\left(0^n + 2^n\right) = \begin{cases} 1, & \text{if } n = 0; \\ 2^{n-1}, & \text{if } n > 0 \end{cases}$.

Proof of Lemma 7.30. Each of the two claims of Lemma 7.30 follows by a straightforward case distinction (which we leave to the reader). □

We can now give answers to parts **(b)**, **(c)**, **(d)** and **(e)** of Exercise 3.18:

Proposition 7.31. Let $a \in \mathbb{N}$. The binomial transform of the list $\left(\binom{0}{a}, \binom{1}{a}, \ldots, \binom{N}{a}\right)$ is the list

$$\left((-1)^a [0 = a], (-1)^a [1 = a], \ldots, (-1)^a [N = a]\right).$$

[320]

Proof of Proposition 7.31. Let $(b_0, b_1, \ldots, b_N)$ be the binomial transform of the list $\left(\binom{0}{a}, \binom{1}{a}, \ldots, \binom{N}{a}\right)$. Thus,

$$b_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{i}{a} \qquad \text{for each } n \in \{0, 1, \ldots, N\}$$

(by the definition of the binomial transform).

Hence, each $n \in \{0, 1, \ldots, N\}$ satisfies

$$\begin{aligned} b_n &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{i}{a} = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \binom{j}{a} \\ &\qquad \left(\text{here, we have renamed the summation index } i \text{ as } j\right) \\ &= (-1)^a [n = a] \qquad \left(\text{by Corollary 7.29 (applied to } i = a)\right). \end{aligned}$$

In other words,

$$(b_0, b_1, \ldots, b_N) = \left((-1)^a [0 = a], (-1)^a [1 = a], \ldots, (-1)^a [N = a]\right). \tag{598}$$

[320] This list $\left((-1)^a [0 = a], (-1)^a [1 = a], \ldots, (-1)^a [N = a]\right)$ actually has a very simple form:

- If it has an $(a+1)$-st entry (i.e., if $a \leq N$), then this entry is $(-1)^a \underbrace{[a = a]}_{\substack{=1 \\ (\text{since } a=a)}} = (-1)^a$.
- All other entries are 0 (because if $n \in \{0, 1, \ldots, N\}$ is such that $n \neq a$, then $(-1)^a \underbrace{[n = a]}_{\substack{=0 \\ (\text{since } n \neq a)}} = 0$).

Thus, this list $\left((-1)^a [0 = a], (-1)^a [1 = a], \ldots, (-1)^a [N = a]\right)$ can be rewritten as $\left(0, 0, \ldots, 0, (-1)^a, 0, 0, \ldots, 0\right)$ (with the $(-1)^a$ entry being placed in the $(a+1)$-st position) when $a \leq N$, and as $(0, 0, \ldots, 0)$ if $a > N$.

But recall that the binomial transform of the list $\left(\binom{0}{a}, \binom{1}{a}, \ldots, \binom{N}{a}\right)$ is $(b_0, b_1, \ldots, b_N)$. In view of (598), this rewrites as follows: The binomial transform of the list $\left(\binom{0}{a}, \binom{1}{a}, \ldots, \binom{N}{a}\right)$ is $\left((-1)^a [0 = a], (-1)^a [1 = a], \ldots, (-1)^a [N = a]\right)$. This proves Proposition 7.31. $\square$

Proposition 7.32. The binomial transform of the list $(1, 1, \ldots, 1)$ (with $N + 1$ entries) is the list $(1, 0, 0, \ldots, 0)$ (with one 1 and N zeroes).

Proof of Proposition 7.32. We have $\binom{n}{0} = 1$ for each $n \in \{0, 1, \ldots, N\}$ (by (227) (applied to $m = n$)). Hence,

$$\left(\binom{0}{0}, \binom{1}{0}, \ldots, \binom{N}{0}\right) = (1, 1, \ldots, 1) \tag{599}$$

(with $N + 1$ entries).

But Proposition 7.31 (applied to $a = 0$) shows that the binomial transform of the list $\left(\binom{0}{0}, \binom{1}{0}, \ldots, \binom{N}{0}\right)$ is the list

$$\left(\underbrace{(-1)^0}_{=1} [0 = 0], \underbrace{(-1)^0}_{=1} [1 = 0], \ldots, \underbrace{(-1)^0}_{=1} [N = 0]\right)$$
$$= ([0 = 0], [1 = 0], \ldots, [N = 0]) = (1, 0, 0, \ldots, 0)$$

(with one 1 and N zeroes). In view of (599), this rewrites as follows: The binomial transform of the list $(1, 1, \ldots, 1)$ (with $N + 1$ entries) is the list $(1, 0, 0, \ldots, 0)$ (with one 1 and N zeroes). This proves Proposition 7.32. $\square$

Proposition 7.33. Let $q \in \mathbb{Z}$. The binomial transform of the list $\left(q^0, q^1, \ldots, q^N\right)$ is $\left((1 - q)^0, (1 - q)^1, \ldots, (1 - q)^N\right)$.

Proof of Proposition 7.33. Let $(b_0, b_1, \ldots, b_N)$ be the binomial transform of the list $\left(q^0, q^1, \ldots, q^N\right)$. Thus,

$$b_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} q^i \qquad \text{for each } n \in \{0, 1, \ldots, N\}$$

(by the definition of the binomial transform). Hence, each $n \in \{0, 1, \ldots, N\}$ satisfies

$$b_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} q^i = (1 - q)^n$$

(by Corollary 7.28). In other words,

$$(b_0, b_1, \ldots, b_N) = \left((1-q)^0, (1-q)^1, \ldots, (1-q)^N\right). \qquad (600)$$

Recall that the binomial transform of the list $\left(q^0, q^1, \ldots, q^N\right)$ is $(b_0, b_1, \ldots, b_N)$. In view of (600), this rewrites as follows: The binomial transform of the list $\left(q^0, q^1, \ldots, q^N\right)$ is $\left((1-q)^0, (1-q)^1, \ldots, (1-q)^N\right)$. This proves Proposition 7.33. □

Proposition 7.34. The binomial transform of the list $(1, 0, 1, 0, 1, 0, \ldots)$ (with $N+1$ entries) is $\left(1, 2^0, 2^1, \ldots, 2^{N-1}\right)$.

Proof of Proposition 7.34. Let $(a_0, a_1, \ldots, a_N)$ be the list $(1, 0, 1, 0, 1, 0, \ldots)$ (with $N+1$ entries). Thus, for each $i \in \{0, 1, \ldots, N\}$, we have

$$a_i = \begin{cases} 1, & \text{if } i \text{ is even;} \\ 0, & \text{if } i \text{ is odd} \end{cases}. \qquad (601)$$

Each $i \in \mathbb{N}$ satisfies

$$\frac{1}{2}\left(1 + (-1)^i\right) = \begin{cases} 1, & \text{if } i \text{ is even;} \\ 0, & \text{if } i \text{ is odd} \end{cases} \qquad (602)$$

(by Lemma 7.30 **(a)** (applied to $n = i$)). Hence, each $i \in \{0, 1, \ldots, N\}$ satisfies

$$\begin{aligned} a_i &= \begin{cases} 1, & \text{if } i \text{ is even;} \\ 0, & \text{if } i \text{ is odd} \end{cases} && \text{(by (601))} \\ &= \frac{1}{2}\left(1 + (-1)^i\right) && \text{(by (602))}. \end{aligned} \qquad (603)$$

Let $(b_0, b_1, \ldots, b_N)$ be the binomial transform of the list $(1, 0, 1, 0, 1, 0, \ldots)$ (with $N+1$ entries). In other words, $(b_0, b_1, \ldots, b_N)$ is the binomial transform of the list $(a_0, a_1, \ldots, a_N)$ (because the list $(1, 0, 1, 0, 1, 0, \ldots)$ (with $N+1$ entries) is precisely $(a_0, a_1, \ldots, a_N)$). Hence,

$$b_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} a_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}$$

(by the definition of the binomial transform). Thus, for each $n \in \{0, 1, \ldots, N\}$, we

obtain

$$\begin{aligned}
b_n &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \underbrace{a_i}_{\substack{=\frac{1}{2}\left(1+(-1)^i\right) \\ \text{(by (603))}}} = \sum_{i=0}^{n} \underbrace{(-1)^i \binom{n}{i} \cdot \frac{1}{2}\left(1+(-1)^i\right)}_{\substack{=\frac{1}{2}(-1)^i \binom{n}{i}\left(1+(-1)^i\right) \\ =\frac{1}{2}(-1)^i \binom{n}{i}\cdot 1+\frac{1}{2}(-1)^i \binom{n}{i}(-1)^i}} \\
&= \sum_{i=0}^{n} \left(\frac{1}{2} (-1)^i \binom{n}{i} \cdot 1 + \frac{1}{2} (-1)^i \binom{n}{i} (-1)^i \right) \\
&= \frac{1}{2} \sum_{i=0}^{n} (-1)^i \binom{n}{i} \cdot \underbrace{1}_{=1^i} + \frac{1}{2} \sum_{i=0}^{n} (-1)^i \binom{n}{i} (-1)^i \\
&= \frac{1}{2} \underbrace{\sum_{i=0}^{n} (-1)^i \binom{n}{i} \cdot 1^i}_{\substack{=(1-1)^n \\ \text{(by Corollary 7.28, applied to } q=1)}} + \frac{1}{2} \underbrace{\sum_{i=0}^{n} (-1)^i \binom{n}{i} (-1)^i}_{\substack{=(1-(-1))^n \\ \text{(by Corollary 7.28, applied to } q=-1)}} \\
&= \frac{1}{2} \Big(\underbrace{1-1}_{=0} \Big)^n + \frac{1}{2} \Big(\underbrace{1-(-1)}_{=2} \Big)^n = \frac{1}{2} 0^n + \frac{1}{2} 2^n \\
&= \frac{1}{2} \left(0^n + 2^n \right) = \begin{cases} 1, & \text{if } n = 0; \\ 2^{n-1}, & \text{if } n > 0 \end{cases}
\end{aligned}$$

(by Lemma 7.30 **(b)**). In other words,

$$(b_0, b_1, \ldots, b_N) = \left(1, 2^0, 2^1, \ldots, 2^{N-1}\right). \tag{604}$$

Recall that the binomial transform of the list $(1, 0, 1, 0, 1, 0, \ldots)$ (with $N + 1$ entries) is $(b_0, b_1, \ldots, b_N)$. In view of (604), this rewrites as follows: The binomial transform of the list $(1, 0, 1, 0, 1, 0, \ldots)$ (with $N + 1$ entries) is $\left(1, 2^0, 2^1, \ldots, 2^{N-1}\right)$. This proves Proposition 7.34. □

Next, we shall arm ourselves with another elementary lemma:

Lemma 7.35. Let N, n and j be nonnegative integers such that $N \geq n$ and $N \geq j$. Then,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j} = \binom{N-n}{N-j}.$$

There are two ways to prove Lemma 7.35. One way is combinatorial (using the principle of inclusion and exclusion) and is explained in [Galvin17, proof of Identity 17.1]. The other way is algebraic; this is the way we shall now present.

Proof of Lemma 7.35. We have $j \leq N$ (since $N \geq j$). Thus, Proposition 3.32 **(e)** (applied to j, n and N instead of x, y and n) yields

$$\binom{n-j-1}{N-j} = \sum_{k=0}^{N} (-1)^{k-j} \binom{k}{j} \binom{n}{N-k}. \qquad (605)$$

But $N - j \geq 0$ (since $N \geq j$), so that $N - j \in \mathbb{N}$. Hence, Proposition 3.16 (applied to $N - n$ and $N - j$ instead of m and n) yields

$$\begin{aligned}
\binom{N-n}{N-j} &= (-1)^{N-j} \underbrace{\binom{(N-j)-(N-n)-1}{N-j}}_{\substack{=\binom{n-j-1}{N-j} \\ (\text{since } (N-j)-(N-n)-1=n-j-1)}} \\
&= (-1)^{N-j} \underbrace{\binom{n-j-1}{N-j}}_{\substack{=\sum_{k=0}^{N} (-1)^{k-j} \binom{k}{j} \binom{n}{N-k} \\ (\text{by (605)})}} = (-1)^{N-j} \sum_{k=0}^{N} (-1)^{k-j} \binom{k}{j} \binom{n}{N-k} \\
&= \sum_{k=0}^{N} \underbrace{(-1)^{N-j} (-1)^{k-j}}_{\substack{=(-1)^{(N-j)+(k-j)}=(-1)^{N-k} \\ (\text{since } (N-j)+(k-j)=N+k-2j\equiv N+k\equiv N-k \operatorname{mod} 2)}} \binom{k}{j} \binom{n}{N-k} \\
&= \sum_{k=0}^{N} (-1)^{N-k} \binom{k}{j} \binom{n}{N-k}. \qquad (606)
\end{aligned}$$

If i is an integer satisfying $i \geq n + 1$, then

$$\binom{n}{i} = 0 \qquad (607)$$

[321].

But $0 \leq n \leq N$ (since $N \geq n$). Hence, we can split the sum $\sum_{i=0}^{N} (-1)^{i} \binom{n}{i} \binom{N-i}{j}$

[321] *Proof of (607):* Let i be an integer satisfying $i \geq n + 1$. Thus, $i \geq n + 1 > n \geq 0$, so that $i \in \mathbb{N}$. Also, from $i > n$, we obtain $n < i$. Hence, Proposition 3.6 (applied to n and i instead of m and n) yields $\binom{n}{i} = 0$. This proves (607).

at $i = n$. We thus find

$$\begin{aligned}
&\sum_{i=0}^{N} (-1)^i \binom{n}{i} \binom{N-i}{j} \\
&= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j} + \sum_{i=n+1}^{N} (-1)^i \underbrace{\binom{n}{i}}_{\substack{=0 \\ \text{(by (607))}}} \binom{N-i}{j} \\
&= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j} + \underbrace{\sum_{i=n+1}^{N} (-1)^i 0 \binom{N-i}{j}}_{=0} \\
&= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j} \\
&= \sum_{i=0}^{N} (-1)^i \binom{n}{i} \binom{N-i}{j} = \sum_{k=0}^{N} (-1)^{N-k} \binom{n}{N-k} \underbrace{\binom{N-(N-k)}{j}}_{\substack{=\binom{k}{j} \\ \text{(since } N-(N-k)=k)}} \\
&\qquad \text{(here, we have substituted } N-k \text{ for } i \text{ in the sum)} \\
&= \sum_{k=0}^{N} (-1)^{N-k} \underbrace{\binom{n}{N-k} \binom{k}{j}}_{=\binom{k}{j}\binom{n}{N-k}} = \sum_{k=0}^{N} (-1)^{N-k} \binom{k}{j} \binom{n}{N-k} = \binom{N-n}{N-j}
\end{aligned}$$

(by (606)). This proves Lemma 7.35. □

Let us now make two definitions that are slight variations on the definitions of B and W made in Exercise 3.18:

> **Definition 7.36. (a)** Let $B : \mathbb{Q}^{N+1} \to \mathbb{Q}^{N+1}$ be the map which sends every list $(f_0, f_1, \ldots, f_N) \in \mathbb{Q}^{N+1}$ to its binomial transform $(g_0, g_1, \ldots, g_N) \in \mathbb{Q}^{N+1}$.
>
> **(b)** Let $W : \mathbb{Q}^{N+1} \to \mathbb{Q}^{N+1}$ be the map which sends every list $(f_0, f_1, \ldots, f_N) \in \mathbb{Q}^{N+1}$ to the list $\left((-1)^N f_N, (-1)^N f_{N-1}, \ldots, (-1)^N f_0\right) \in \mathbb{Q}^{N+1}$.

The maps B and W we have just defined are **almost** the same as the maps B and W from Exercise 3.18 **(f)**. The only difference is that the former maps are defined on

$\mathbb{Q}^{N+1}$ (and have codomains $\mathbb{Q}^{N+1}$ as well), whereas the latter are defined on $\mathbb{Z}^{N+1}$ (and have codomains $\mathbb{Z}^{N+1}$). Thus, the latter maps are restrictions of the former maps. Hence, in order to prove that the latter maps satisfy the two equalities $B \circ W \circ B = W \circ B \circ W$ and $(B \circ W)^3 = \text{id}$ (as demanded by Exercise 3.18 **(f)**), it is perfectly sufficient to show that the former maps satisfy these two equalities. In other words, it is perfectly sufficient to prove the following theorem:

Theorem 7.37. The two maps B and W introduced in Definition 7.36 satisfy $B \circ W \circ B = W \circ B \circ W$ and $(B \circ W)^3 = \text{id}$.

Before we prove Theorem 7.37, let us show three simpler facts:

Lemma 7.38. Let $\mathbf{f} = (f_0, f_1, \ldots, f_N)$ and $\mathbf{g} = (g_0, g_1, \ldots, g_N)$ be two lists such that $\mathbf{g} = B(\mathbf{f})$.

(a) Then,

$$g_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} f_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}.$$

(b) Also,

$$g_n = \sum_{i=0}^{N} (-1)^i \binom{n}{i} f_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}.$$

Proof of Lemma 7.38. **(a)** We know that the map B sends the list $(f_0, f_1, \ldots, f_N)$ to its binomial transform (by the definition of B). In other words, $B((f_0, f_1, \ldots, f_N))$ is the binomial transform of $(f_0, f_1, \ldots, f_N)$. In other words, $(g_0, g_1, \ldots, g_N)$ is the binomial transform of $(f_0, f_1, \ldots, f_N)$ (since $B\left(\underbrace{(f_0, f_1, \ldots, f_N)}_{=\mathbf{f}}\right) = B(\mathbf{f}) = \mathbf{g} = (g_0, g_1, \ldots, g_N)$). In other words, we have

$$g_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} f_i \qquad \text{for each } n \in \{0, 1, \ldots, N\}$$

(by the definition of the binomial transform). This proves Lemma 7.38 **(a)**.

(b) Let $n \in \{0, 1, \ldots, N\}$. Thus, $0 \leq n \leq N$.

Lemma 7.38 **(a)** yields

$$g_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} f_i. \tag{608}$$

For each integer i satisfying $i \geq n+1$, we have

$$\binom{n}{i} = 0 \tag{609}$$

[322].

But $0 \leq n \leq N$. Hence, we can split the sum $\sum_{i=0}^{N} (-1)^i \dbinom{n}{i} f_i$ at $i = n$. We thus find

$$\sum_{i=0}^{N} (-1)^i \dbinom{n}{i} f_i = \sum_{i=0}^{n} (-1)^i \dbinom{n}{i} f_i + \sum_{i=n+1}^{N} (-1)^i \underbrace{\dbinom{n}{i}}_{\substack{=0 \\ \text{(by (609))}}} f_i$$
$$= \sum_{i=0}^{n} (-1)^i \dbinom{n}{i} f_i + \underbrace{\sum_{i=n+1}^{N} (-1)^i 0 f_i}_{=0} = \sum_{i=0}^{n} (-1)^i \dbinom{n}{i} f_i = g_n$$

(by (608)). In other words, $g_n = \sum_{i=0}^{N} (-1)^i \dbinom{n}{i} f_i$. This proves Lemma 7.38 **(b)**. □

Proposition 7.39. The map B introduced in Definition 7.36 satisfies $B \circ B = \operatorname{id}$.

Proof of Proposition 7.39. Let $\mathbf{a}$ be a list. Write this list $\mathbf{a}$ in the form $\mathbf{a} = (a_0, a_1, \ldots, a_N)$.

Write the list $B(\mathbf{a})$ in the form $B(\mathbf{a}) = (b_0, b_1, \ldots, b_N)$.

We know that the map B sends the list $(a_0, a_1, \ldots, a_N)$ to its binomial transform (by the definition of B). In other words, $B((a_0, a_1, \ldots, a_N))$ is the binomial transform of $(a_0, a_1, \ldots, a_N)$. In other words, $(b_0, b_1, \ldots, b_N)$ is the binomial transform of $(a_0, a_1, \ldots, a_N)$ (since $B\left(\underbrace{(a_0, a_1, \ldots, a_N)}_{=\mathbf{a}}\right) = B(\mathbf{a}) = (b_0, b_1, \ldots, b_N)$). Hence, Corollary 7.27 (applied to $f_i = a_i$ and $g_i = b_i$) shows that $(a_0, a_1, \ldots, a_N)$ is the binomial transform of $(b_0, b_1, \ldots, b_N)$.

We know that the map B sends the list $(b_0, b_1, \ldots, b_N)$ to its binomial transform (by the definition of B). In other words, $B((b_0, b_1, \ldots, b_N))$ is the binomial transform of $(b_0, b_1, \ldots, b_N)$.

The two lists $B((b_0, b_1, \ldots, b_N))$ and $(a_0, a_1, \ldots, a_N)$ must be equal, since each of them is the binomial transform of $(b_0, b_1, \ldots, b_N)$ (as we have proven above). In other words, $B((b_0, b_1, \ldots, b_N)) = (a_0, a_1, \ldots, a_N)$. Now,

$$(B \circ B)(\mathbf{a}) = B\left(\underbrace{B(\mathbf{a})}_{=(b_0, b_1, \ldots, b_N)}\right) = B((b_0, b_1, \ldots, b_N)) = (a_0, a_1, \ldots, a_N)$$
$$= \mathbf{a} = \operatorname{id}(\mathbf{a}).$$

Now, forget that we fixed $\mathbf{a}$. We thus have shown that $(B \circ B)(\mathbf{a}) = \operatorname{id}(\mathbf{a})$ for each list $\mathbf{a}$. In other words, $B \circ B = \operatorname{id}$. This proves Proposition 7.39. □

[322] *Proof of (609):* Let i be an integer satisfying $i \geq n+1$. We have $i \geq n+1 > n \geq 0$, so that $i \in \mathbb{N}$. Also, $n \in \{0, 1, \ldots, N\} \subseteq \mathbb{N}$. Also, $n < i$ (since $i > n$). Hence, Proposition 3.6 (applied to n and i instead of m and n) yields $\dbinom{n}{i} = 0$. This proves (609).

Proposition 7.40. The map W introduced in Definition 7.36 satisfies $W \circ W = \operatorname{id}$.

Proof of Proposition 7.40. Let $\mathbf{a}$ be a list. Write $\mathbf{a}$ in the form $\mathbf{a} = (a_0, a_1, \ldots, a_N)$.

The map W sends the list $(a_0, a_1, \ldots, a_N)$ to $\left((-1)^N a_N, (-1)^N a_{N-1}, \ldots, (-1)^N a_0\right)$ (by the definition of W). In other words,

$$W((a_0, a_1, \ldots, a_N)) = \left((-1)^N a_N, (-1)^N a_{N-1}, \ldots, (-1)^N a_0\right).$$

Let $\mathbf{b}$ be the list $W(\mathbf{a})$. Thus, $\mathbf{b} = W(\mathbf{a})$. Write the list $\mathbf{b}$ in the form $\mathbf{b} = (b_0, b_1, \ldots, b_N)$. Thus,

$$\begin{aligned}(b_0, b_1, \ldots, b_N) = \mathbf{b} = W\left(\underbrace{\mathbf{a}}_{=(a_0,a_1,\ldots,a_N)}\right) \\ = W((a_0, a_1, \ldots, a_N)) = \left((-1)^N a_N, (-1)^N a_{N-1}, \ldots, (-1)^N a_0\right).\end{aligned}$$

In other words,

$$b_n = (-1)^N a_{N-n} \qquad \text{for each } n \in \{0, 1, \ldots, N\}. \tag{610}$$

Hence, for each $n \in \{0, 1, \ldots, N\}$, we have

$$\begin{aligned}(-1)^N \underbrace{b_{N-n}}_{\substack{=(-1)^N a_{N-(N-n)} \\ \text{(by (610) (applied} \\ \text{to } N-n \text{ instead of } n))}} &= \underbrace{(-1)^N (-1)^N}_{=((-1)(-1))^N} \underbrace{a_{N-(N-n)}}_{\substack{=a_n \\ (\text{since } N-(N-n)=n)}} \\ &= \left(\underbrace{(-1)(-1)}_{=1}\right)^N a_n = \underbrace{1^N}_{=1} a_n = a_n.\end{aligned}$$

In other words,

$$\left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right) = (a_0, a_1, \ldots, a_N).$$

But the map W sends the list $(b_0, b_1, \ldots, b_N)$ to $\left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right)$ (by the definition of W). In other words,

$$W((b_0, b_1, \ldots, b_N)) = \left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right).$$

Hence,

$$\begin{aligned}(W \circ W)(\mathbf{a}) = W\left(\underbrace{W(\mathbf{a})}_{=\mathbf{b}=(b_0,b_1,\ldots,b_N)}\right) = W((b_0, b_1, \ldots, b_N)) \\ = \left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right) = (a_0, a_1, \ldots, a_N) \\ = \mathbf{a} = \operatorname{id}(\mathbf{a}).\end{aligned}$$

Now, forget that we fixed $\mathbf{a}$. We thus have shown that $(W \circ W)(\mathbf{a}) = \operatorname{id}(\mathbf{a})$ for each list $\mathbf{a}$. In other words, $W \circ W = \operatorname{id}$. This proves Proposition 7.40. $\square$

We are now ready to prove Theorem 7.37:

Proof of Theorem 7.37. Let us first focus on proving that $B \circ W \circ B = W \circ B \circ W$.

Indeed, let $\mathbf{a}$ be a list. Write $\mathbf{a}$ in the form $\mathbf{a} = (a_0, a_1, \ldots, a_N)$.

Let $\mathbf{b}$ be the list $B(\mathbf{a})$. Thus, $\mathbf{b} = B(\mathbf{a})$. Write the list $\mathbf{b}$ in the form $\mathbf{b} = (b_0, b_1, \ldots, b_N)$.

Lemma 7.38 **(b)** (applied to $\mathbf{a}$, a_i, $\mathbf{b}$ and b_i instead of $\mathbf{f}$, f_i, $\mathbf{g}$ and g_i) shows that

$$b_n = \sum_{i=0}^{N} (-1)^i \binom{n}{i} a_i \qquad \text{for each } n \in \{0, 1, \ldots, N\} \tag{611}$$

(since $\mathbf{b} = B(\mathbf{a})$).

The map W sends the list $(b_0, b_1, \ldots, b_N)$ to $\left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right)$ (by the definition of W). In other words,

$$W((b_0, b_1, \ldots, b_N)) = \left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right).$$

Thus,

$$W\left(\underbrace{\mathbf{b}}_{=(b_0, b_1, \ldots, b_N)}\right) = W((b_0, b_1, \ldots, b_N)) = \left((-1)^N b_N, (-1)^N b_{N-1}, \ldots, (-1)^N b_0\right). \tag{612}$$

Now, define a list $\mathbf{c}$ by $\mathbf{c} = B(W(\mathbf{b}))$. Write the list $\mathbf{c}$ in the form $\mathbf{c} = (c_0, c_1, \ldots, c_N)$. Hence, Lemma 7.38 **(a)** (applied to $W(\mathbf{b})$, $(-1)^N b_{N-i}$, $\mathbf{c}$ and c_i instead of $\mathbf{f}$, f_i, $\mathbf{g}$ and g_i) shows that

$$c_n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} (-1)^N b_{N-i} \qquad \text{for each } n \in \{0, 1, \ldots, N\} \tag{613}$$

(because $\mathbf{c} = B(W(\mathbf{b}))$ and because of (612)).

The map W sends the list $(a_0, a_1, \ldots, a_N)$ to $\left((-1)^N a_N, (-1)^N a_{N-1}, \ldots, (-1)^N a_0\right)$ (by the definition of W). In other words,

$$W((a_0, a_1, \ldots, a_N)) = \left((-1)^N a_N, (-1)^N a_{N-1}, \ldots, (-1)^N a_0\right).$$

Thus,

$$W\left(\underbrace{\mathbf{a}}_{=(a_0, a_1, \ldots, a_N)}\right) = W((a_0, a_1, \ldots, a_N)) = \left((-1)^N a_N, (-1)^N a_{N-1}, \ldots, (-1)^N a_0\right). \tag{614}$$

Define a list $\mathbf{d}$ by $\mathbf{d} = B(W(\mathbf{a}))$. Write the list $\mathbf{d}$ in the form $\mathbf{d} = (d_0, d_1, \ldots, d_N)$. Hence, Lemma 7.38 **(b)** (applied to $W(\mathbf{a})$, $(-1)^N a_{N-i}$, $\mathbf{d}$ and d_i instead of $\mathbf{f}$, f_i, $\mathbf{g}$ and g_i) shows that

$$d_n = \sum_{i=0}^{N} (-1)^i \binom{n}{i} (-1)^N a_{N-i} \qquad \text{for each } n \in \{0, 1, \ldots, N\} \tag{615}$$

(because $\mathbf{d} = B(W(\mathbf{a}))$ and because of (614)).

The map W sends the list $(d_0, d_1, \ldots, d_N)$ to $\left((-1)^N d_N, (-1)^N d_{N-1}, \ldots, (-1)^N d_0\right)$ (by the definition of W). In other words,

$$W((d_0, d_1, \ldots, d_N)) = \left((-1)^N d_N, (-1)^N d_{N-1}, \ldots, (-1)^N d_0\right).$$

Thus,

$$\begin{aligned} W\left(\underbrace{\mathbf{d}}_{=(d_0,d_1,\ldots,d_N)}\right) &= W((d_0, d_1, \ldots, d_N)) \\ &= \left((-1)^N d_N, (-1)^N d_{N-1}, \ldots, (-1)^N d_0\right). \end{aligned} \tag{616}$$

We shall now show that $\mathbf{c} = W(\mathbf{d})$.

Indeed, for any $g \in \{0, 1, \ldots, N\}$, we have

$$\begin{aligned} b_g &= \sum_{i=0}^{N} (-1)^i \binom{g}{i} a_i \qquad \text{(by (611), applied to } n = g\text{)} \\ &= \sum_{j=0}^{N} (-1)^j \binom{g}{j} a_j \end{aligned} \tag{617}$$

(here, we have renamed the summation index i as j).

Now, let $n \in \{0, 1, \ldots, N\}$ be arbitrary. Then, $n \leq N$, so that $N \geq n$. Hence, $N - n \geq 0$, so that $0 \leq N - n \leq N$ (since $n \geq 0$ (since $n \in \{0, 1, \ldots, N\}$)). Thus, $N - n \in \{0, 1, \ldots, N\}$. Hence, (615) (applied to $N - n$ instead of n) yields

$$d_{N-n} = \sum_{i=0}^{N} (-1)^i \binom{N-n}{i} (-1)^N a_{N-i}.$$

Multiplying both sides of this equality by $(-1)^N$, we obtain

$$\begin{aligned}
(-1)^N d_{N-n} &= (-1)^N \left(\sum_{i=0}^{N} (-1)^i \binom{N-n}{i} (-1)^N a_{N-i} \right) \\
&= \sum_{i=0}^{N} (-1)^i \binom{N-n}{i} \underbrace{(-1)^N (-1)^N}_{\substack{=((-1)(-1))^N = 1^N \\ (\text{since } (-1)(-1)=1)}} a_{N-i} \\
&= \sum_{i=0}^{N} (-1)^i \binom{N-n}{i} \underbrace{1^N}_{=1} a_{N-i} \\
&= \sum_{i=0}^{N} (-1)^i \binom{N-n}{i} a_{N-i}. \qquad (618)
\end{aligned}$$

For each $i \in \{0, 1, \ldots, n\}$, we have

$$b_{N-i} = \sum_{j=0}^{N} (-1)^j \binom{N-i}{j} a_j \qquad (619)$$

[323].

[323] *Proof of (619):* Let $i \in \{0, 1, \ldots, n\}$. Thus, $i \in \{0, 1, \ldots, n\} \subseteq \{0, 1, \ldots, N\}$ (since $n \leq N$). Hence, $N - i \in \{0, 1, \ldots, N\}$. Thus, (617) (applied to $g = N - i$) yields $b_{N-i} = \sum_{j=0}^{N} (-1)^j \binom{N-i}{j} a_j$. This proves (619).

But (613) becomes

$$\begin{aligned}
c_n &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} (-1)^N \underbrace{b_{N-i}}_{\substack{=\sum_{j=0}^{N} (-1)^j \binom{N-i}{j} a_j \\ \text{(by (619))}}} \\
&= \sum_{i=0}^{n} (-1)^i \binom{n}{i} (-1)^N \sum_{j=0}^{N} (-1)^j \binom{N-i}{j} a_j \\
&= \underbrace{\sum_{i=0}^{n} \sum_{j=0}^{N}}_{=\sum_{j=0}^{N} \sum_{i=0}^{n}} (-1)^i \binom{n}{i} (-1)^N (-1)^j \binom{N-i}{j} a_j \\
&= \sum_{j=0}^{N} \sum_{i=0}^{n} (-1)^i \binom{n}{i} (-1)^N (-1)^j \binom{N-i}{j} a_j \\
&= \sum_{j=0}^{N} \underbrace{(-1)^N (-1)^j}_{\substack{=(-1)^{N+j}=(-1)^{N-j} \\ \text{(since } N+j \equiv N-j \bmod 2)}} \underbrace{\left(\sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j} \right)}_{\substack{=\binom{N-n}{N-j} \\ \text{(by Lemma 7.35 (since } N \geq j \\ \text{(since } j \leq N)))}} \underbrace{a_j}_{\substack{=a_{N-(N-j)} \\ \text{(since } j=N-(N-j))}} \\
&= \sum_{j=0}^{N} (-1)^{N-j} \binom{N-n}{N-j} a_{N-(N-j)} = \sum_{i=0}^{N} (-1)^i \binom{N-n}{i} a_{N-i} \\
&\qquad \left(\text{here, we have substituted } i \text{ for } N-j \text{ in the sum}\right) \\
&= (-1)^N d_{N-n} \qquad \text{(by (618))}.
\end{aligned}$$

Now, forget that we fixed n. We thus have proven that $c_n = (-1)^N d_{N-n}$ for each $n \in \{0, 1, \ldots, N\}$. In other words,

$$(c_0, c_1, \ldots, c_N) = \left((-1)^N d_N, (-1)^N d_{N-1}, \ldots, (-1)^N d_0 \right).$$

Thus,

$$\begin{aligned}
(B\circ W\circ B)(\mathbf{a}) &= B\left(W\left(\underbrace{B(\mathbf{a})}_{=\mathbf{b}}\right)\right) = B(W(\mathbf{b})) = \mathbf{c} \qquad (\text{since } \mathbf{c} = B(W(\mathbf{b}))) \\
&= (c_0, c_1, \ldots, c_N) = \left((-1)^N d_N, (-1)^N d_{N-1}, \ldots, (-1)^N d_0\right) \\
&= W\left(\underbrace{\mathbf{d}}_{=B(W(\mathbf{a}))}\right) \qquad (\text{by (616)}) \\
&= W(B(W(\mathbf{a}))) = (W\circ B\circ W)(\mathbf{a}).
\end{aligned}$$

Now, forget that we fixed $\mathbf{a}$. We thus have proven that $(B\circ W\circ B)(\mathbf{a}) = (W\circ B\circ W)(\mathbf{a})$ for each list $\mathbf{a}$. In other words,

$$B\circ W\circ B = W\circ B\circ W. \tag{620}$$

Hence,

$$\begin{aligned}
(B\circ W)^3 &= (B\circ W)\circ(B\circ W)\circ(B\circ W) \\
&= \underbrace{B\circ W\circ B}_{\substack{=W\circ B\circ W \\ (\text{by (620)})}} \circ W\circ B\circ W = W\circ B\circ \underbrace{W\circ W}_{\substack{=\text{id} \\ (\text{by Proposition 7.40})}} \circ B\circ W \\
&= W\circ \underbrace{B\circ B}_{\substack{=\text{id} \\ (\text{by Proposition 7.39})}} \circ W = W\circ W = \text{id} \qquad (\text{by Proposition 7.40}).
\end{aligned}$$

This completes the proof of Theorem 7.37. □

Solution to Exercise 3.18. Recall that the definition of the binomial transform in Definition 7.23 generalizes the definition of the binomial transform we gave in Exercise 3.18. Hence, part **(a)** of Exercise 3.18 follows from Corollary 7.27. Part **(b)** follows from Proposition 7.32. Part **(c)** follows from Proposition 7.31. Part **(d)** follows from Proposition 7.33. Part **(e)** follows from Proposition 7.34. Finally, part **(f)** follows from Theorem 7.37 (since the maps B and W from Exercise 3.18 **(f)** are restrictions of the maps B and W from Definition 7.36). Hence, Exercise 3.18 is solved. □

7.27. Solution to Exercise 3.19

Before we solve Exercise 3.19, we state a lemma:

Lemma 7.41. Let $n \in \mathbb{N}$. Then,

$$\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{k} \binom{n+1}{k} = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \frac{1}{n+1}.$$

Proof of Lemma 7.41. Clearly, $n+1$ is a positive integer (since $n \in \mathbb{N}$). In other words, $n+1 \in \{1,2,3,\ldots\}$. Hence, $n+1 \neq 0$. Thus, the fraction $\frac{1}{n+1}$ is well-defined.

Also, for each $k \in \{1,2,3,\ldots\}$, the fraction $\frac{(-1)^{k-1}}{k}$ is well-defined (since $k \neq 0$).

Now, let $k \in \{1,2,3,\ldots\}$. Then, Proposition 3.22 (applied to $n+1$ and k instead of m and n) yields $\binom{n+1}{k} = \frac{n+1}{k}\binom{(n+1)-1}{k-1} = \frac{n+1}{k}\binom{n}{k-1}$ (since $(n+1)-1 = n$). Dividing both sides of this equality by $n+1$, we find

$$\frac{1}{n+1}\binom{n+1}{k} = \frac{1}{n+1}\cdot\frac{n+1}{k}\binom{n}{k-1} = \frac{1}{k}\binom{n}{k-1}. \tag{621}$$

Now,

$$\begin{aligned}\frac{(-1)^{k-1}}{k}\binom{n}{k-1} &= \underbrace{(-1)^{k-1}}_{=-(-1)^k}\cdot\underbrace{\frac{1}{k}\binom{n}{k-1}}_{\substack{=\frac{1}{n+1}\binom{n+1}{k}\\ \text{(by (621))}}} = -(-1)^k\cdot\frac{1}{n+1}\binom{n+1}{k}\\ &= \frac{-1}{n+1}(-1)^k\binom{n+1}{k}.\end{aligned} \tag{622}$$

Now, forget that we fixed k. We thus have proven the equality (622) for each $k \in \{1,2,3,\ldots\}$.

On the other hand, Proposition 3.39 **(c)** (applied to $n+1$ instead of n) yields

$$\sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k} = \begin{cases}1, & \text{if } n+1=0;\\ 0, & \text{if } n+1\neq 0\end{cases} = 0$$

(since $n+1 \neq 0$). Hence,

$$\begin{aligned}0 &= \sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k} = \underbrace{(-1)^0}_{=1}\underbrace{\binom{n+1}{0}}_{\substack{=1\\ \text{(by (227) (applied to } m=n+1))}} + \sum_{k=1}^{n+1}(-1)^k\binom{n+1}{k}\\ &\qquad\text{(here, we have split off the addend for } k=0 \text{ from the sum)}\\ &= 1+\sum_{k=1}^{n+1}(-1)^k\binom{n+1}{k}.\end{aligned}$$

Subtracting 1 from this equality, we obtain

$$-1 = \sum_{k=1}^{n+1}(-1)^k\binom{n+1}{k}. \tag{623}$$

Also, $n < n+1$. Hence, Proposition 3.6 (applied to n and $n+1$ instead of m and n) yields

$$\binom{n}{n+1} = 0.$$

Now,

$$\begin{aligned}
\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{k} \binom{n}{k} &= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \frac{(-1)^{(n+1)-1}}{n+1} \underbrace{\binom{n}{n+1}}_{=0} \\
&\qquad \text{(here, we have split off the addend for } k = n+1 \text{ from the sum)} \\
&= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \underbrace{\frac{(-1)^{(n+1)-1}}{n+1} 0}_{=0} \\
&= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k}.
\end{aligned} \tag{624}$$

Finally, each $k \in \{1, 2, 3, \ldots\}$ satisfies

$$\begin{aligned}
\binom{n+1}{k} &= \binom{(n+1)-1}{k-1} + \binom{(n+1)-1}{k} \\
&\qquad \left(\begin{array}{c} \text{by Proposition 3.11} \\ \text{(applied to } n+1 \text{ and } k \text{ instead of } m \text{ and } n) \end{array} \right) \\
&= \binom{n}{k-1} + \binom{n}{k} \qquad \left(\text{since } (n+1)-1 = n\right) \\
&= \binom{n}{k} + \binom{n}{k-1}.
\end{aligned} \tag{625}$$

Now,

$$\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{k} \underbrace{\binom{n+1}{k}}_{\substack{=\binom{n}{k}+\binom{n}{k-1} \\ \text{(by (625))}}}$$
$$= \sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{k} \left(\binom{n}{k} + \binom{n}{k-1} \right) = \underbrace{\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{k} \binom{n}{k}}_{\substack{= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} \\ \text{(by (624))}}} + \sum_{k=1}^{n+1} \underbrace{\frac{(-1)^{k-1}}{k} \binom{n}{k-1}}_{\substack{= \frac{-1}{n+1} (-1)^k \binom{n+1}{k} \\ \text{(by (622))}}}$$
$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \underbrace{\sum_{k=1}^{n+1} \frac{-1}{n+1} (-1)^k \binom{n+1}{k}}_{= \frac{-1}{n+1} \sum_{k=1}^{n+1} (-1)^k \binom{n+1}{k}}$$
$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \frac{-1}{n+1} \underbrace{\sum_{k=1}^{n+1} (-1)^k \binom{n+1}{k}}_{\substack{=-1 \\ \text{(by (623))}}}$$
$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \frac{-1}{n+1} (-1) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} + \frac{1}{n+1}.$$

This proves Lemma 7.41. □

Solution to Exercise 3.19. We shall solve Exercise 3.19 by induction on n:

Induction base: Comparing $\sum_{k=1}^{0} \frac{(-1)^{k-1}}{k} \binom{0}{k} = (\text{empty sum}) = 0$ with $\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{0} = (\text{empty sum}) = 0$, we obtain $\sum_{k=1}^{0} \frac{(-1)^{k-1}}{k} \binom{0}{k} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{0}$. In other words, Exercise 3.19 holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Exercise 3.19 holds for $n = m$. We must prove that Exercise 3.19 holds for $n = m + 1$.

We have assumed that Exercise 3.19 holds for $n = m$. In other words, we have

$$\sum_{k=1}^{m} \frac{(-1)^{k-1}}{k} \binom{m}{k} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{m}.$$

Now, Lemma 7.41 (applied to $n = m$) yields

$$\sum_{k=1}^{m+1} \frac{(-1)^{k-1}}{k} \binom{m+1}{k} = \underbrace{\sum_{k=1}^{m} \frac{(-1)^{k-1}}{k} \binom{m}{k}}_{=\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{m}} + \frac{1}{m+1}$$
$$= \left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{m}\right) + \frac{1}{m+1} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{m+1}.$$

In other words, Exercise 3.19 holds for $n = m + 1$. This completes the induction step. Thus, Exercise 3.19 is solved. □

7.28. Solution to Exercise 3.20

We begin with a few lemmas before we come to the solution of Exercise 3.20. The first lemma is essentially trivial:

Lemma 7.42. Let $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$. Then, $\binom{n}{k}$ is a positive integer.

Proof of Lemma 7.42. From $n \in \mathbb{N} \subseteq \mathbb{Z}$ and $k \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$, we conclude that $\binom{n}{k} \in \mathbb{Z}$ (by Proposition 3.20 (applied to n and k instead of m and n)). In other words, $\binom{n}{k}$ is an integer.

From $k \in \{0, 1, \ldots, n\}$, we obtain $k \leq n$, so that $n \geq k$. Also, $k \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$. Hence, Proposition 3.4 (applied to n and k instead of m and n) yields $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$. But $\frac{n!}{k!\,(n-k)!}$ is a positive rational number (since $n!$, $k!$ and $(n-k)!$ are positive integers). In other words, $\binom{n}{k}$ is a positive rational number (since $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$). Hence, $\binom{n}{k}$ is a positive integer (since $\binom{n}{k}$ is an integer). This proves Lemma 7.42. □

Lemma 7.43. Let $q \in \mathbb{Q}$ and $m \in \mathbb{N}$ be such that $q \neq m$. Then,

$$\binom{q}{m} = \frac{q}{q-m} \binom{q-1}{m}.$$

Proof of Lemma 7.43. We have $q - m \neq 0$ (since $q \neq m$). Hence, the fraction $\frac{q}{q-m}$ is well-defined.

The equality (226) (applied to q and m instead of m and n) yields

$$\binom{q}{m} = \frac{q(q-1)\cdots(q-m+1)}{m!}. \tag{626}$$

The equality (226) (applied to $q-1$ and m instead of m and n) yields

$$\begin{aligned}\binom{q-1}{m} &= \frac{(q-1)((q-1)-1)\cdots((q-1)-m+1)}{m!} \\ &= \frac{1}{m!}\cdot\underbrace{((q-1)((q-1)-1)\cdots((q-1)-m+1))}_{\substack{=(q-1)(q-2)\cdots(q-m) \\ \text{(since } (q-1)-1=q-2 \text{ and } (q-1)-m+1=q-m)}} \\ &= \frac{1}{m!}\cdot((q-1)(q-2)\cdots(q-m)).\end{aligned}$$

Multiplying both sides of this equality by $\frac{q}{q-m}$, we find

$$\begin{aligned}\frac{q}{q-m}\binom{q-1}{m} &= \frac{q}{q-m}\cdot\frac{1}{m!}\cdot((q-1)(q-2)\cdots(q-m)) \\ &= \frac{1}{q-m}\cdot\frac{1}{m!}\cdot\underbrace{q\cdot((q-1)(q-2)\cdots(q-m))}_{\substack{=q(q-1)\cdots(q-m) \\ =(q(q-1)\cdots(q-m+1))\cdot(q-m)}} \\ &= \frac{1}{q-m}\cdot\frac{1}{m!}\cdot(q(q-1)\cdots(q-m+1))\cdot(q-m) \\ &= \frac{1}{m!}\cdot(q(q-1)\cdots(q-m+1)) = \frac{q(q-1)\cdots(q-m+1)}{m!}.\end{aligned}$$

Comparing this with (626), we obtain $\binom{q}{m} = \frac{q}{q-m}\binom{q-1}{m}$. This proves Lemma 7.43. □

Lemma 7.44. Let $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$. Then:

(a) We have

$$\binom{n+1}{k+1} = \frac{n+1}{k+1}\binom{n}{k}.$$

(b) We have

$$\binom{n+1}{k} = \frac{n+1}{n+1-k}\binom{n}{k}.$$

Proof of Lemma 7.44. We have $k \in \{0, 1, \ldots, n\}$, thus $k \leq n < n+1$. Hence, $n+1-k > 0$. Thus, the fraction $\frac{n+1}{n+1-k}$ is well-defined.

Also, from $k \in \{0, 1, \ldots, n\}$, we obtain $k + 1 \in \{1, 2, \ldots, n + 1\}$; thus, $k + 1 > 0$. Hence, the fraction $\dfrac{n+1}{k+1}$ is well-defined.

We have $k \in \{0, 1, \ldots, n\}$, so that $k + 1 \in \{1, 2, \ldots, n + 1\} \subseteq \{1, 2, 3, \ldots\}$. Thus, Proposition 3.22 (applied to $n + 1$ and $k + 1$ instead of m and n) yields

$$\binom{n+1}{k+1} = \frac{n+1}{k+1}\binom{(n+1)-1}{(k+1)-1} = \frac{n+1}{k+1}\binom{n}{k} \tag{627}$$

(since $(k+1) - 1 = k$ and $(n+1) - 1 = n$). This proves Lemma 7.44 **(a)**.

(b) We have $k < n + 1$ and thus $k \neq n + 1$, so that $n + 1 \neq k$. Also, $n + 1 \in \mathbb{N} \subseteq \mathbb{Q}$. Hence, Lemma 7.43 (applied to $q = n + 1$ and $m = k$) yields

$$\binom{n+1}{k} = \frac{n+1}{n+1-k}\binom{n+1-1}{k} = \frac{n+1}{n+1-k}\binom{n}{k}$$

(since $n + 1 - 1 = n$). This proves Lemma 7.44 **(b)**. $\square$

Our next lemma is a simple consequence of the recurrence of the binomial coefficients:

Lemma 7.45. Let $n \in \mathbb{N}$. Let $k \in \{0, 1, \ldots, n\}$. Then,

$$\frac{1}{\binom{n}{k}} = \left(\frac{1}{\binom{n+1}{k}} + \frac{1}{\binom{n+1}{k+1}}\right)\frac{n+1}{n+2}. \tag{628}$$

(In particular, all the fractions $\dfrac{1}{\binom{n}{k}}$, $\dfrac{1}{\binom{n+1}{k}}$ and $\dfrac{1}{\binom{n+1}{k+1}}$ in this equality are well-defined.)

Proof of Lemma 7.45. Lemma 7.42 shows that $\binom{n}{k}$ is a positive integer. Thus, $\binom{n}{k} \neq 0$. Hence, the fraction $\dfrac{1}{\binom{n}{k}}$ is well-defined.

We have $k \in \{0, 1, \ldots, n\} \subseteq \{0, 1, \ldots, n + 1\}$. Thus, Lemma 7.42 (applied to $n + 1$ instead of n) shows that $\binom{n+1}{k}$ is a positive integer. Thus, $\binom{n+1}{k} \neq 0$. Hence, the fraction $\dfrac{1}{\binom{n+1}{k}}$ is well-defined.

We have $k \in \{0,1,\ldots,n\}$, thus $k+1 \in \{1,2,\ldots,n+1\} \subseteq \{0,1,\ldots,n+1\}$. Thus, Lemma 7.42 (applied to $n+1$ and $k+1$ instead of n and k) shows that $\dbinom{n+1}{k+1}$ is a positive integer. Thus, $\dbinom{n+1}{k+1} \neq 0$. Hence, the fraction $\dfrac{1}{\dbinom{n+1}{k+1}}$ is well-defined.

It remains to prove the equality (628). We have

$$\begin{aligned}
&\frac{1}{\dbinom{n+1}{k}} + \frac{1}{\dbinom{n+1}{k+1}} \\
&= 1/ \underbrace{\dbinom{n+1}{k}}_{\substack{=\frac{n+1}{n+1-k}\binom{n}{k} \\ \text{(by Lemma 7.44 (b))}}} + 1/ \underbrace{\dbinom{n+1}{k+1}}_{\substack{=\frac{n+1}{k+1}\binom{n}{k} \\ \text{(by Lemma 7.44 (a))}}} \\
&= \underbrace{1/\left(\frac{n+1}{n+1-k}\binom{n}{k}\right)}_{=\frac{n+1-k}{n+1}/\binom{n}{k}} + \underbrace{1/\left(\frac{n+1}{k+1}\binom{n}{k}\right)}_{=\frac{k+1}{n+1}/\binom{n}{k}} \\
&= \frac{n+1-k}{n+1}/\binom{n}{k} + \frac{k+1}{n+1}/\binom{n}{k} = \underbrace{\left(\frac{n+1-k}{n+1} + \frac{k+1}{n+1}\right)}_{=\frac{n+2}{n+1}}/\binom{n}{k} \\
&= \frac{n+2}{n+1}/\binom{n}{k}.
\end{aligned}$$

Multiplying both sides of this equality by $\dfrac{n+1}{n+2}$, we obtain

$$\left(\frac{1}{\dbinom{n+1}{k}} + \frac{1}{\dbinom{n+1}{k+1}}\right)\frac{n+1}{n+2} = \left(\frac{n+2}{n+1}/\binom{n}{k}\right)\frac{n+1}{n+2} = \frac{1}{\dbinom{n}{k}}.$$

This proves Lemma 7.45. $\square$

We can now prove part **(a)** of Exercise 3.20:

Proposition 7.46. Let $n \in \mathbb{N}$.

(a) We have

$$\sum_{k=0}^{n} \frac{(-1)^k}{\dbinom{n}{k}} = \frac{n+1}{n+2}\left(1+(-1)^n\right).$$

(b) We have
$$\sum_{k=0}^{n} \frac{(-1)^k}{\dbinom{n}{k}} = 2 \cdot \frac{n+1}{n+2} [n \text{ is even}] .$$
(Here, we are using the Iverson bracket notation, as in Definition 3.48; thus, $[n \text{ is even}]$ is 1 if n is even and 0 otherwise.)

Proof of Proposition 7.46. **(a)** We have $\dbinom{n+1}{0} = 1$ (by Proposition 3.3 **(a)** (applied to $m = n+1$)), thus $\dfrac{1}{\dbinom{n+1}{0}} = \dfrac{1}{1} = 1$.

Also, $\dbinom{n+1}{n+1} = 1$ (by Proposition 3.9 (applied to $m = n+1$)), thus $\dfrac{1}{\dbinom{n+1}{n+1}} = \dfrac{1}{1} = 1$.

We have

$$\sum_{k=0}^{n} \underbrace{\frac{(-1)^k}{\dbinom{n}{k}}}_{=(-1)^k \cdot \frac{1}{\binom{n}{k}}} = \sum_{k=0}^{n} (-1)^k \cdot \underbrace{\frac{1}{\dbinom{n}{k}}}_{\substack{= \left(\frac{1}{\binom{n+1}{k}} + \frac{1}{\binom{n+1}{k+1}} \right) \frac{n+1}{n+2} \\ \text{(by (628))}}}$$
$$= \sum_{k=0}^{n} (-1)^k \cdot \left(\frac{1}{\dbinom{n+1}{k}} + \frac{1}{\dbinom{n+1}{k+1}} \right) \frac{n+1}{n+2}$$
$$= \frac{n+1}{n+2} \sum_{k=0}^{n} (-1)^k \cdot \left(\frac{1}{\dbinom{n+1}{k}} + \frac{1}{\dbinom{n+1}{k+1}} \right) . \qquad (629)$$

But

$$\sum_{k=0}^{n} \underbrace{(-1)^k \cdot \left(\frac{1}{\dbinom{n+1}{k}} + \frac{1}{\dbinom{n+1}{k+1}} \right)}_{=(-1)^k \cdot \frac{1}{\binom{n+1}{k}} + (-1)^k \cdot \frac{1}{\binom{n+1}{k+1}}}$$

$$= \sum_{k=0}^{n} \left((-1)^k \cdot \frac{1}{\dbinom{n+1}{k}} + (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}} \right)$$

$$= \underbrace{\sum_{k=0}^{n} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k}}}_{\substack{=(-1)^0 \cdot \frac{1}{\binom{n+1}{0}} + \sum_{k=1}^{n} (-1)^k \cdot \frac{1}{\binom{n+1}{k}} \\ \text{(here, we have split off the addend} \\ \text{for } k=0 \text{ from the sum)}}} + \underbrace{\sum_{k=0}^{n} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}}}_{\substack{=(-1)^n \cdot \frac{1}{\binom{n+1}{n+1}} + \sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\binom{n+1}{k+1}} \\ \text{(here, we have split off the addend} \\ \text{for } k=n \text{ from the sum)}}}$$

$$= \underbrace{(-1)^0}_{=1} \cdot \frac{1}{\underbrace{\dbinom{n+1}{0}}_{=1}} + \underbrace{\sum_{k=1}^{n} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k}}}_{\substack{=\sum_{k=0}^{n-1} (-1)^{k+1} \cdot \frac{1}{\binom{n+1}{k+1}} \\ \text{(here, we have substituted } k+1 \\ \text{for } k \text{ in the sum)}}} + (-1)^n \cdot \frac{1}{\underbrace{\dbinom{n+1}{n+1}}_{=1}} + \sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}}$$

$$= 1 + \sum_{k=0}^{n-1} \underbrace{(-1)^{k+1}}_{=-(-1)^k} \cdot \frac{1}{\dbinom{n+1}{k+1}} + (-1)^n + \sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}}$$

$$= 1 + \underbrace{\sum_{k=0}^{n-1} \left(-(-1)^k \right) \cdot \frac{1}{\dbinom{n+1}{k+1}}}_{=-\sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\binom{n+1}{k+1}}} + (-1)^n + \sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}}$$

$$= 1 + \left(-\sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}} \right) + (-1)^n + \sum_{k=0}^{n-1} (-1)^k \cdot \frac{1}{\dbinom{n+1}{k+1}} = 1 + (-1)^n .$$

Hence, (629) becomes

$$\sum_{k=0}^{n} \frac{(-1)^k}{\dbinom{n}{k}} = \frac{n+1}{n+2} \underbrace{\sum_{k=0}^{n} (-1)^k \cdot \left(\frac{1}{\dbinom{n+1}{k}} + \frac{1}{\dbinom{n+1}{k+1}} \right)}_{=1+(-1)^n}$$
$$= \frac{n+1}{n+2} \left(1 + (-1)^n\right).$$

This proves Proposition 7.46 **(a)**.

(b) The definition of the truth value $[n \text{ is even}]$ shows that

$$[n \text{ is even}] = \begin{cases} 1, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is not even} \end{cases} = \begin{cases} 1, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \qquad (630)$$

(because the condition "n is not even" is equivalent to "n is odd"). But Proposition 7.46 **(a)** yields

$$\sum_{k=0}^{n} \frac{(-1)^k}{\dbinom{n}{k}} = \frac{n+1}{n+2} \left(1 + (-1)^n\right) = 2 \cdot \frac{n+1}{n+2} \cdot \underbrace{\frac{1}{2} \left(1 + (-1)^n\right)}_{\substack{= \begin{cases} 1, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \\ \text{(by Lemma 7.30 (a))}}}$$
$$= 2 \cdot \frac{n+1}{n+2} \cdot \underbrace{\begin{cases} 1, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}}_{\substack{=[n \text{ is even}] \\ \text{(by (630))}}} = 2 \cdot \frac{n+1}{n+2} [n \text{ is even}].$$

This proves Proposition 7.46 **(b)**. □

We remark that Proposition 7.46 **(a)** can also be seen as a particular case of the identity

$$\sum_{k=0}^{n} \frac{(-1)^k}{\dbinom{x}{k}} = \left(1 + \frac{(-1)^n}{\dbinom{x+1}{n+1}} \right) \frac{x+1}{x+2}$$

(for all $n \in \{-1, 0, 1, \ldots\}$ and $x \in \mathbb{Q} \setminus \{-2, -1, 0, \ldots, n-1\}$), which was observed by user "user90369" on `https://math.stackexchange.com/a/3251880/` and can be proven fairly easily by induction on n.

Solution to Exercise 3.20. **(a)** Proposition 7.46 **(b)** yields

$$\sum_{k=0}^{n} \frac{(-1)^k}{\dbinom{n}{k}} = 2 \cdot \frac{n+1}{n+2} [n \text{ is even}].$$

This solves Exercise 3.20 **(a)**.

(b) Let us forget that we fixed n. We shall solve Exercise 3.20 **(b)** by induction on n:

Induction base: Proposition 3.3 **(a)** (applied to $m = 0$) yields $\binom{0}{0} = 1$. We have $0 + 1 = 1$ and thus

$$\frac{0+1}{2^{0+1}} \sum_{k=1}^{0+1} \frac{2^k}{k} = \frac{1}{2^1} \underbrace{\sum_{k=1}^{1} \frac{2^k}{k}}_{= \frac{2^1}{1}} = \frac{1}{2^1} \cdot \frac{2^1}{1} = 1.$$

Comparing this with

$$\sum_{k=0}^{0} \frac{1}{\binom{0}{k}} = \frac{1}{\binom{0}{0}} = 1/\underbrace{\binom{0}{0}}_{=1} = 1/1 = 1,$$

we obtain $\sum_{k=0}^{0} \frac{1}{\binom{0}{k}} = \frac{0+1}{2^{0+1}} \sum_{k=1}^{0+1} \frac{2^k}{k}$. In other words, Exercise 3.20 **(b)** holds for $n = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Exercise 3.20 **(b)** holds for $n = m$. We must prove that Exercise 3.20 **(b)** holds for $n = m + 1$.

We have assumed that Exercise 3.20 **(b)** holds for $n = m$. In other words, we have

$$\sum_{k=0}^{m} \frac{1}{\binom{m}{k}} = \frac{m+1}{2^{m+1}} \sum_{k=1}^{m+1} \frac{2^k}{k}. \qquad (631)$$

We have $\binom{m+1}{0} = 1$ (by Proposition 3.3 **(a)** (applied to $m + 1$ instead of m)), thus $\frac{1}{\binom{m+1}{0}} = \frac{1}{1} = 1$.

Also, $\binom{m+1}{m+1} = 1$ (by Proposition 3.9 (applied to $m + 1$ instead of m)), thus $\frac{1}{\binom{m+1}{m+1}} = \frac{1}{1} = 1$.

Notice that $m + 1 > m \geq 0$ (since $m \in \mathbb{N}$), and thus $m + 1 \neq 0$.

We have

$$\sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} = \sum_{k=0}^{m} \frac{1}{\binom{m+1}{k}} + \frac{1}{\underbrace{\binom{m+1}{m+1}}_{=1}}$$
$$\left(\begin{array}{c}\text{here, we have split off the addend for } k=m+1 \\ \text{from the sum}\end{array}\right)$$
$$= \sum_{k=0}^{m} \frac{1}{\binom{m+1}{k}} + 1.$$

Thus,

$$\sum_{k=0}^{m} \frac{1}{\binom{m+1}{k}} = \sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} - 1. \tag{632}$$

Also,

$$\sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} = \frac{1}{\underbrace{\binom{m+1}{0}}_{=1}} + \sum_{k=1}^{m+1} \frac{1}{\binom{m+1}{k}}$$
$$\left(\begin{array}{c}\text{here, we have split off the addend for } k=0 \\ \text{from the sum}\end{array}\right)$$
$$= 1 + \sum_{k=1}^{m+1} \frac{1}{\binom{m+1}{k}} = 1 + \sum_{k=0}^{m} \frac{1}{\binom{m+1}{k+1}}$$

(here, we have substituted $k+1$ for k in the sum). Thus,

$$\sum_{k=0}^{m} \frac{1}{\binom{m+1}{k+1}} = \sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} - 1. \tag{633}$$

Every $k \in \{0, 1, \ldots, m\}$ satisfies

$$\frac{1}{\binom{m}{k}} = \left(\frac{1}{\binom{m+1}{k}} + \frac{1}{\binom{m+1}{k+1}} \right) \frac{m+1}{m+2} \tag{634}$$

(by (628) (applied to $n = m$)).

Now,

$$\sum_{k=0}^{m} \underbrace{\frac{1}{\dbinom{m}{k}}}_{= \left(\frac{1}{\binom{m+1}{k}} + \frac{1}{\binom{m+1}{k+1}} \right) \frac{m+1}{m+2}} = \sum_{k=0}^{m} \left(\frac{1}{\dbinom{m+1}{k}} + \frac{1}{\dbinom{m+1}{k+1}} \right) \frac{m+1}{m+2}$$

$$= \frac{m+1}{m+2} \sum_{k=0}^{m} \left(\frac{1}{\dbinom{m+1}{k}} + \frac{1}{\dbinom{m+1}{k+1}} \right).$$

In view of

$$\sum_{k=0}^{m} \left(\frac{1}{\dbinom{m+1}{k}} + \frac{1}{\dbinom{m+1}{k+1}} \right) = \underbrace{\sum_{k=0}^{m} \frac{1}{\dbinom{m+1}{k}}}_{\substack{= \sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} - 1 \\ \text{(by (632))}}} + \underbrace{\sum_{k=0}^{m} \frac{1}{\dbinom{m+1}{k+1}}}_{\substack{= \sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} - 1 \\ \text{(by (633))}}}$$

$$= \left(\sum_{k=0}^{m+1} \frac{1}{\dbinom{m+1}{k}} - 1 \right) + \left(\sum_{k=0}^{m+1} \frac{1}{\dbinom{m+1}{k}} - 1 \right)$$

$$= 2 \left(\sum_{k=0}^{m+1} \frac{1}{\dbinom{m+1}{k}} - 1 \right),$$

this becomes

$$\sum_{k=0}^{m} \frac{1}{\dbinom{m}{k}} = \frac{m+1}{m+2} \underbrace{\sum_{k=0}^{m} \left(\frac{1}{\dbinom{m+1}{k}} + \frac{1}{\dbinom{m+1}{k+1}} \right)}_{= 2 \left(\sum_{k=0}^{m+1} \frac{1}{\binom{m+1}{k}} - 1 \right)} = \frac{m+1}{m+2} \cdot 2 \left(\sum_{k=0}^{m+1} \frac{1}{\dbinom{m+1}{k}} - 1 \right).$$

Comparing this with (631), we obtain

$$\frac{m+1}{m+2}\cdot 2\left(\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}-1\right)=\frac{m+1}{2^{m+1}}\sum_{k=1}^{m+1}\frac{2^k}{k}.$$

Multiplying both sides of this equality by $\frac{m+2}{m+1}$ (this is allowed, since $m+1\neq 0$), we obtain

$$\frac{m+2}{m+1}\cdot\frac{m+1}{m+2}\cdot 2\left(\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}-1\right)=\frac{m+2}{m+1}\cdot\frac{m+1}{2^{m+1}}\sum_{k=1}^{m+1}\frac{2^k}{k}=\frac{m+2}{2^{m+1}}\sum_{k=1}^{m+1}\frac{2^k}{k}.$$

Hence,

$$\frac{m+2}{2^{m+1}}\sum_{k=1}^{m+1}\frac{2^k}{k}=\frac{m+2}{m+1}\cdot\frac{m+1}{m+2}\cdot 2\left(\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}-1\right)=2\left(\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}-1\right).$$

Dividing both sides of this equality by 2, we find

$$\frac{1}{2}\cdot\frac{m+2}{2^{m+1}}\sum_{k=1}^{m+1}\frac{2^k}{k}=\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}-1.$$

Solving this equality for $\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}$, we obtain

$$\sum_{k=0}^{m+1}\frac{1}{\dbinom{m+1}{k}}=\underbrace{\frac{1}{2}\cdot\frac{m+2}{2^{m+1}}}_{\substack{=\frac{m+2}{2\cdot 2^{m+1}}=\frac{m+2}{2^{m+2}}\\ \text{(since } 2\cdot 2^{m+1}=2^{(m+1)+1}=2^{m+2}\\ \text{(because } (m+1)+1=m+2\text{))}}}\sum_{k=1}^{m+1}\frac{2^k}{k}+1=\frac{m+2}{2^{m+2}}\sum_{k=1}^{m+1}\frac{2^k}{k}+1.$$

Comparing this with

$$\underbrace{\frac{(m+1)+1}{2^{(m+1)+1}}}_{\substack{=\frac{m+2}{2^{m+2}} \\ \text{(since } (m+1)+1=m+2)}} \underbrace{\sum_{k=1}^{(m+1)+1} \frac{2^k}{k}}_{\substack{=\sum_{k=1}^{m+2} \frac{2^k}{k} \\ \text{(since } (m+1)+1=m+2)}}$$
$$= \frac{m+2}{2^{m+2}} \underbrace{\sum_{k=1}^{m+2} \frac{2^k}{k}}_{\substack{=\sum_{k=1}^{m+1} \frac{2^k}{k} + \frac{2^{m+2}}{m+2} \\ \text{(here, we have split off the} \\ \text{addend for } k=m+2 \text{ from the sum)}}} = \frac{m+2}{2^{m+2}} \left(\sum_{k=1}^{m+1} \frac{2^k}{k} + \frac{2^{m+2}}{m+2} \right)$$
$$= \frac{m+2}{2^{m+2}} \sum_{k=1}^{m+1} \frac{2^k}{k} + \underbrace{\frac{m+2}{2^{m+2}} \cdot \frac{2^{m+2}}{m+2}}_{=1} = \frac{m+2}{2^{m+2}} \sum_{k=1}^{m+1} \frac{2^k}{k} + 1,$$

we obtain

$$\sum_{k=0}^{m+1} \frac{1}{\dbinom{m+1}{k}} = \frac{(m+1)+1}{2^{(m+1)+1}} \sum_{k=1}^{(m+1)+1} \frac{2^k}{k}.$$

In other words, Exercise 3.20 **(b)** holds for $n = m+1$. This completes the induction step. Thus, Exercise 3.20 **(b)** is solved by induction. □

7.29. Solution to Exercise 3.21

We shall prove the following generalization of Exercise 3.21:

Proposition 7.47. Let $\mathbb{K}$ be a commutative ring. (See Definition 6.2 for the definition of a "commutative ring". For example, we can set $\mathbb{K} = \mathbb{Z}$ or $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{Q}[X]$.) Let x and y be two elements of $\mathbb{K}$. For any $n \in \mathbb{N}$ and $m \in \mathbb{N}$, define $Y_{m,n} \in \mathbb{K}$ by

$$Y_{m,n} = \sum_{k=0}^{n} y^k \binom{n}{k} \left(x^{n-k} + y\right)^m .$$

Then, $Y_{m,n} = Y_{n,m}$ for any $n \in \mathbb{N}$ and $m \in \mathbb{N}$.

Proof of Proposition 7.47. For any $n \in \mathbb{N}$ and $m \in \mathbb{N}$, we have

$$\begin{aligned}
Y_{m,n} &= \sum_{k=0}^{n} y^k \binom{n}{k} \left(x^{n-k} + y\right)^m = \sum_{\ell=0}^{n} y^\ell \binom{n}{\ell} \Bigg(\underbrace{x^{n-\ell} + y}_{=y+x^{n-\ell}} \Bigg)^m \\
&\qquad \text{(here, we have renamed the summation index } k \text{ as } \ell\text{)} \\
&= \sum_{\ell=0}^{n} y^\ell \binom{n}{\ell} \underbrace{\left(y + x^{n-\ell}\right)^m}_{\substack{=\sum_{k=0}^{m} \binom{m}{k} y^k \left(x^{n-\ell}\right)^{m-k} \\ \text{(by the binomial formula)}}} = \sum_{\ell=0}^{n} y^\ell \binom{n}{\ell} \left(\sum_{k=0}^{m} \binom{m}{k} y^k \left(x^{n-\ell}\right)^{m-k} \right) \\
&= \sum_{\ell=0}^{n} \sum_{k=0}^{m} \underbrace{y^\ell \binom{n}{\ell} \binom{m}{k}}_{=\binom{n}{\ell}\binom{m}{k} y^\ell} y^k \left(x^{n-\ell}\right)^{m-k} = \sum_{\ell=0}^{n} \sum_{k=0}^{m} \binom{n}{\ell} \binom{m}{k} \underbrace{y^\ell y^k}_{=y^{\ell+k}} \underbrace{\left(x^{n-\ell}\right)^{m-k}}_{=x^{(n-\ell)(m-k)}} \\
&= \sum_{\ell=0}^{n} \sum_{k=0}^{m} \binom{n}{\ell} \binom{m}{k} y^{\ell+k} x^{(n-\ell)(m-k)}. \qquad (635)
\end{aligned}$$

Now, let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then, (635) (applied to m and n instead of n and m) shows that

$$\begin{aligned}
Y_{n,m} &= \underbrace{\sum_{\ell=0}^{m} \sum_{k=0}^{n}}_{=\sum_{k=0}^{n} \sum_{\ell=0}^{m}} \underbrace{\binom{m}{\ell} \binom{n}{k}}_{=\binom{n}{k}\binom{m}{\ell}} \underbrace{y^{\ell+k}}_{=y^{k+\ell}} \underbrace{x^{(m-\ell)(n-k)}}_{=x^{(n-k)(m-\ell)}} \\
&= \sum_{k=0}^{n} \sum_{\ell=0}^{m} \binom{n}{k} \binom{m}{\ell} y^{k+\ell} x^{(n-k)(m-\ell)} = \sum_{k=0}^{n} \sum_{g=0}^{m} \binom{n}{k} \binom{m}{g} y^{k+g} x^{(n-k)(m-g)} \\
&\qquad \text{(here, we have renamed the summation index } \ell \text{ as } g\text{)} \\
&= \sum_{\ell=0}^{n} \sum_{g=0}^{m} \binom{n}{\ell} \binom{m}{g} y^{\ell+g} x^{(n-\ell)(m-g)} \\
&\qquad \text{(here, we have renamed the summation index } k \text{ as } \ell\text{)} \\
&= \sum_{\ell=0}^{n} \sum_{k=0}^{m} \binom{n}{\ell} \binom{m}{k} y^{\ell+k} x^{(n-\ell)(m-k)} \\
&\qquad \text{(here, we have renamed the summation index } g \text{ as } k\text{)} \\
&= Y_{m,n} \qquad \text{(by (635))}.
\end{aligned}$$

This proves Proposition 7.47. □

Solution to Exercise 3.21. Set $\mathbb{K} = \mathbb{Z}[X]$, and define two elements x and y of $\mathbb{K}$ by $x = X$ and $y = -1$. For any $n \in \mathbb{N}$ and $m \in \mathbb{N}$, define $Y_{m,n} \in \mathbb{K}$ as in Proposition

7.47. Then, for any $n \in \mathbb{N}$ and $m \in \mathbb{N}$, we have

$$Y_{m,n} = \sum_{k=0}^{n} \underbrace{y^k}_{\substack{=(-1)^k \\ (\text{since } y=-1)}} \binom{n}{k} \left(\underbrace{x^{n-k}}_{\substack{=X^{n-k} \\ (\text{since } x=X)}} + \underbrace{y}_{=-1} \right)^m$$
$$= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(\underbrace{X^{n-k} + (-1)}_{=X^{n-k}-1} \right)^m$$
$$= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(X^{n-k} - 1 \right)^m = Z_{m,n}. \qquad (636)$$

Now, fix $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Applying (636) to m and n instead of n and m, we obtain $Y_{n,m} = Z_{n,m}$. But Proposition 7.47 shows that $Y_{m,n} = Y_{n,m}$. Comparing this with (636), we obtain $Y_{n,m} = Z_{m,n}$. Comparing this with $Y_{n,m} = Z_{n,m}$, we obtain $Z_{m,n} = Z_{n,m}$. This solves Exercise 3.21. □

Remark 7.48. Two solutions to Exercise 3.21 are sketched in `http://www.artofproblemsolving.com/community/c6h333199p1782800` . One is essentially the solution given above (except in lesser generality); the other is combinatorial.

7.30. Solution to Exercise 3.22

We shall give two solutions to Exercise 3.22. The first solution follows the Hint given in the exercise, and illustrates both the use of Lemma 3.28 **(b)** and of "generating functions" (the strategy of proving identities by identifying both sides as coefficients in polynomials or power series). The second solution is of a more classical nature, using no new methods but a tricky application of Theorem 3.30.

7.30.1. First solution

The crux of the first solution is the proof of the following lemma (which appears in [GrKnPa94, (5.55)]):

Lemma 7.49. Let $n \in \mathbb{N}$ and $x \in \mathbb{N}$. Then,

$$\sum_{k=0}^{n} (-1)^k \binom{x}{k} \binom{x}{n-k} = \begin{cases} (-1)^{n/2} \dbinom{x}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

Thus, Lemma 7.49 is obtained from Exercise 3.22 by substituting a nonnegative integer x for the indeterminate X. It thus is clear that Lemma 7.49 follows from

Exercise 3.22. However, for us, the interest lies in the opposite implication: We shall derive Exercise 3.22 from Lemma 7.49. Let us, however, prove Lemma 7.49 first. But before we do this, let us state a version of the binomial theorem:

Proposition 7.50. Let $x \in \mathbb{N}$. Then,

$$(1+X)^x = \sum_{k\in\mathbb{N}} \binom{x}{k} X^k$$

(an equality between polynomials in $\mathbb{Z}[X]$). (The sum $\sum_{k\in\mathbb{N}} \binom{x}{k} X^k$ is an infinite sum, but only finitely many of its addends are nonzero, so it is well-defined.)

Proof of Proposition 7.50. We have

$$\sum_{k\in\mathbb{N}} \binom{x}{k} X^k = \underbrace{\sum_{\substack{k\in\mathbb{N};\\ k\leq x}}}_{=\sum_{k=0}^{x}} \binom{x}{k} X^k + \sum_{\substack{k\in\mathbb{N};\\ k>x}} \underbrace{\binom{x}{k}}_{\substack{=0 \\ \text{(by (231) (applied to } x \text{ and } k \\ \text{instead of } m \text{ and } n\text{) (since } x<k \text{ (since } k>x\text{)))}}} X^k$$
$$= \sum_{k=0}^{x} \binom{x}{k} X^k + \underbrace{\sum_{\substack{k\in\mathbb{N};\\ k>x}} 0X^k}_{=0} = \sum_{k=0}^{x} \binom{x}{k} X^k.$$

Comparing this with

$$\Big(\underbrace{1+X}_{=X+1}\Big)^x = (X+1)^x = \sum_{k=0}^{x} \binom{x}{k} X^k \underbrace{1^{x-k}}_{=1} \qquad \text{(by the binomial formula)}$$
$$= \sum_{k=0}^{x} \binom{x}{k} X^k,$$

we obtain $(1+X)^x = \sum_{k\in\mathbb{N}} \binom{x}{k} X^k$. This proves Proposition 7.50. □

Proof of Lemma 7.49. Proposition 7.50 yields

$$(1+X)^x = \sum_{k\in\mathbb{N}} \binom{x}{k} X^k. \tag{637}$$

Substituting $-X$ for X in this equality, we obtain

$$(1+(-X))^x = \sum_{k\in\mathbb{N}} \binom{x}{k} \underbrace{(-X)^k}_{=(-1)^k X^k} = \sum_{k\in\mathbb{N}} \binom{x}{k} (-1)^k X^k = \sum_{k\in\mathbb{N}} (-1)^k \binom{x}{k} X^k.$$

Since $1 + (-X) = 1 - X$, this rewrites as

$$(1 - X)^x = \sum_{k \in \mathbb{N}} (-1)^k \binom{x}{k} X^k.$$

Multiplying this equality with (637), we obtain

$$\begin{aligned}(1 - X)^x (1 + X)^x &= \left(\sum_{k \in \mathbb{N}} (-1)^k \binom{x}{k} X^k \right) \left(\sum_{k \in \mathbb{N}} \binom{x}{k} X^k \right) \\ &= \sum_{k \in \mathbb{N}} \left(\sum_{i=0}^{k} (-1)^i \binom{x}{i} \binom{x}{k-i} \right) X^k\end{aligned}$$

(according to the definition of the product of two polynomials). Hence,

$$\begin{aligned}&\left(\text{the coefficient of } X^n \text{ in } (1 - X)^x (1 + X)^x\right) \\ &= \sum_{i=0}^{n} (-1)^i \binom{x}{i} \binom{x}{n-i} = \sum_{k=0}^{n} (-1)^k \binom{x}{k} \binom{x}{n-k}\end{aligned} \tag{638}$$

(here, we have renamed the summation index i as k).

On the other hand,

$$\begin{aligned}(1 - X)^x (1 + X)^x &= \Big(\underbrace{(1 - X)(1 + X)}_{=1 - X^2 = 1 + (-X^2)} \Big)^x = \left(1 + \left(-X^2\right)\right)^x = \sum_{k \in \mathbb{N}} \binom{x}{k} \underbrace{\left(-X^2\right)^k}_{=(-1)^k (X^2)^k} \\ &\qquad \left(\text{this follows from substituting } -X^2 \text{ for } X \text{ in (637)}\right) \\ &= \sum_{k \in \mathbb{N}} \underbrace{\binom{x}{k} (-1)^k}_{=(-1)^k \binom{x}{k}} \underbrace{\left(X^2\right)^k}_{=X^{2k}} = \sum_{k \in \mathbb{N}} (-1)^k \binom{x}{k} X^{2k}.\end{aligned}$$

Hence,

$$\left(\text{the coefficient of } X^n \text{ in } (1 - X)^x (1 + X)^x\right) = \begin{cases} (-1)^{n/2} \dbinom{x}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

Comparing this with (638), we obtain

$$\sum_{k=0}^{n} (-1)^k \binom{x}{k} \binom{x}{n-k} = \begin{cases} (-1)^{n/2} \dbinom{x}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

This proves Lemma 7.49. □

First solution to Exercise 3.22. Define two polynomials P and Q (with rational coefficients) by

$$P = \sum_{k=0}^{n} (-1)^k \binom{X}{k} \binom{X}{n-k} \tag{639}$$

and

$$Q = \begin{cases} (-1)^{n/2} \dbinom{X}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}. \tag{640}$$

For every $x \in \mathbb{N}$, we have

$$\begin{aligned} P(x) &= \sum_{k=0}^{n} (-1)^k \binom{x}{k} \binom{x}{n-k} \qquad \text{(by the definition of } P\text{)} \\ &= \begin{cases} (-1)^{n/2} \dbinom{x}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \qquad \text{(by Lemma 7.49)} \\ &= Q(x) \qquad \text{(by the definition of } Q\text{)}. \end{aligned}$$

Hence, Lemma 3.28 **(b)** shows that $P = Q$. In light of (639) and (640), this rewrites as

$$\sum_{k=0}^{n} (-1)^k \binom{X}{k} \binom{X}{n-k} = \begin{cases} (-1)^{n/2} \dbinom{X}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

This solves Exercise 3.22. □

7.30.2. Second solution

Now let us prepare for the second solution to Exercise 3.22. We shall use the Iverson bracket notation introduced in Definition 3.48.

We notice that every $n \in \mathbb{N}$ satisfies

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = [n = 0] \tag{641}$$

(by Proposition 7.24, applied to $m = n$).

Next, we state a simple fact:

Lemma 7.51. Let $m \in \mathbb{N}$ and $i \in \mathbb{N}$. Then,

$$\sum_{k=0}^{m} (-1)^k \binom{k}{i} \binom{m}{k} = (-1)^i [m = i].$$

Proof of Lemma 7.51. If $m < i$, then Lemma 7.51 holds[324]. Hence, for the rest of this proof of Lemma 7.51, we can WLOG assume that we don't have $m < i$. Assume this.

[324] *Proof.* Assume that $m < i$. Then, $m \neq i$. Thus, $[m = i] = 0$.

But every $k \in \{0, 1, \ldots, m\}$ satisfies $k \leq m < i$. Hence, every $k \in \{0, 1, \ldots, m\}$ satisfies $\binom{k}{i} = 0$ (by (231), applied to k and i instead of m and n). Now,

$$\sum_{k=0}^{m} (-1)^k \underbrace{\binom{k}{i}}_{=0} \binom{m}{k} = \sum_{k=0}^{m} (-1)^k 0 \binom{m}{k} = 0 = [m = i]$$

(since $[m = i] = 0$). In other words, Lemma 7.51 holds, qed.

We have $m \geq i$ (since we don't have $m < i$). Hence, $m \geq i \geq 0$, so that

$$\begin{aligned}
&\sum_{k=0}^{m} (-1)^k \underbrace{\binom{k}{i}\binom{m}{k}}_{=\binom{m}{k}\binom{k}{i}} \\
&= \sum_{k=0}^{m} (-1)^k \binom{m}{k}\binom{k}{i} \\
&= \sum_{k=0}^{i-1} (-1)^k \binom{m}{k} \underbrace{\binom{k}{i}}_{\substack{=0 \\ \text{(by (231) (applied to } k \text{ and } i \\ \text{instead of } m \text{ and } n\text{) (since } k<i\text{))}}} + \sum_{k=i}^{m} (-1)^k \underbrace{\binom{m}{k}\binom{k}{i}}_{\substack{=\binom{m}{i}\binom{m-i}{k-i} \\ \text{(by (241) (applied to } k \text{ and } i \\ \text{instead of } i \text{ and } a\text{) (since } k\geq i\text{))}}} \\
&= \underbrace{\sum_{k=0}^{i-1} (-1)^k \binom{m}{k} 0}_{=0} + \sum_{k=i}^{m} (-1)^k \binom{m}{i}\binom{m-i}{k-i} = \sum_{k=i}^{m} (-1)^k \binom{m}{i}\binom{m-i}{k-i} \\
&= \sum_{k=0}^{m-i} \underbrace{(-1)^{k+i}}_{=(-1)^k(-1)^i} \binom{m}{i}\binom{m-i}{k} \\
&\qquad \text{(here, we have substituted } k+i \text{ for } k \text{ in the sum)} \\
&= \sum_{k=0}^{m-i} (-1)^k (-1)^i \binom{m}{i}\binom{m-i}{k} = (-1)^i \binom{m}{i} \underbrace{\sum_{k=0}^{m-i} (-1)^k \binom{m-i}{k}}_{\substack{=[m-i=0] \\ \text{(by (641) (applied to } n=m-i\text{))}}} \\
&= (-1)^i \binom{m}{i} \left[\underbrace{m-i=0}_{\substack{\text{this is equivalent to} \\ m=i}} \right] = (-1)^i \binom{m}{i} [m=i] .
\end{aligned} \tag{642}$$

But it is easy to see that $\dbinom{m}{i}[m=i]=[m=i]$ [325]. Hence, (642) becomes

$$\sum_{k=0}^{m}(-1)^k\dbinom{k}{i}\dbinom{m}{k}=(-1)^i\underbrace{\dbinom{m}{i}[m=i]}_{=[m=i]}=(-1)^i[m=i].$$

This proves Lemma 7.51. □

Here comes one more simple lemma:

Lemma 7.52. Let $n\in\mathbb{N}$. Let $a_0,a_1,\ldots,a_n$ be $n+1$ polynomials in the indeterminate X with rational coefficients (that is, $n+1$ elements of $\mathbb{Q}[X]$). Then,

$$\sum_{i=0}^{n}a_i[n-i=i]=\begin{cases}a_{n/2}, & \text{if } n \text{ is even;}\\ 0, & \text{if } n \text{ is odd}\end{cases}.$$

Proof of Lemma 7.52. We have $n\in\mathbb{N}$, so that $n\geq 0$. For every $i\in\{0,1,\ldots,n\}$, we have

$$\Big[\underbrace{n-i=i}_{\text{this is equivalent to } n=2i}\Big]=\Big[\underbrace{n=2i}_{\text{this is equivalent to } i=n/2}\Big]=[i=n/2]. \qquad (643)$$

Thus,

$$\begin{aligned}
&\underbrace{\sum_{i=0}^{n}}_{=\sum_{i\in\{0,1,\ldots,n\}}}a_i\underbrace{[n-i=i]}_{\substack{=[i=n/2]\\ \text{(by (643))}}}\\
&=\sum_{i\in\{0,1,\ldots,n\}}a_i[i=n/2]=\sum_{\substack{i\in\{0,1,\ldots,n\};\\ i=n/2}}a_i\underbrace{[i=n/2]}_{\substack{=1\\ \text{(since } i=n/2)}}+\sum_{\substack{i\in\{0,1,\ldots,n\};\\ i\neq n/2}}a_i\underbrace{[i=n/2]}_{\substack{=0\\ \text{(since } i\neq n/2)}}\\
&\qquad\left(\text{since every } i\in\{0,1,\ldots,n\} \text{ satisfies either } i=n/2 \text{ or } i\neq n/2 \text{ (but not both)}\right)\\
&=\sum_{\substack{i\in\{0,1,\ldots,n\};\\ i=n/2}}a_i+\underbrace{\sum_{\substack{i\in\{0,1,\ldots,n\};\\ i\neq n/2}}a_i0}_{=0}=\sum_{\substack{i\in\{0,1,\ldots,n\};\\ i=n/2}}a_i. \qquad (644)
\end{aligned}$$

[325] *Proof.* We have $\underbrace{\dbinom{i}{i}}_{=1}[i=i]=[i=i]$. In other words, the equality $\dbinom{m}{i}[m=i]=[m=i]$ holds in the case when $m=i$. Therefore, in order to prove this equality, we only need to consider the case when $m\neq i$. So assume that $m\neq i$. Then, $[m=i]=0$, and thus $\dbinom{m}{i}\underbrace{[m=i]}_{=0}=0=[m=i]$, qed.

We must be in one of the following two cases:

Case 1: The number n is even.

Case 2: The number n is odd.

Let us first consider Case 1. In this case, the number n is even. Hence,

$$\begin{cases} a_{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} = a_{n/2}. \tag{645}$$

On the other hand, $n/2 \in \mathbb{Z}$ (since n is even). Combined with $n/2 \geq 0$ (since $n \geq 0$) and $n/2 \leq n$ (for the same reason), this shows that $n/2 \in \{0, 1, \ldots, n\}$. Now, (644) becomes

$$\begin{aligned} \sum_{i=0}^{n} a_i [n - i = i] &= \sum_{\substack{i \in \{0,1,\ldots,n\}; \\ i = n/2}} a_i = a_{n/2} \qquad (\text{since } n/2 \in \{0, 1, \ldots, n\}) \\ &= \begin{cases} a_{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \qquad (\text{by (645)}). \end{aligned}$$

Thus, Lemma 7.52 is proven in Case 1.

Let us now consider Case 2. In this case, the number n is odd. Hence,

$$\begin{cases} a_{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} = 0. \tag{646}$$

On the other hand, $n/2 \notin \mathbb{Z}$ (since n is odd). Hence, $n/2 \notin \{0, 1, \ldots, n\}$ (since $\{0, 1, \ldots, n\} \subseteq \mathbb{Z}$). Now, (644) becomes

$$\begin{aligned} \sum_{i=0}^{n} a_i [n - i = i] &= \sum_{\substack{i \in \{0,1,\ldots,n\}; \\ i = n/2}} a_i = (\text{empty sum}) \qquad (\text{since } n/2 \notin \{0, 1, \ldots, n\}) \\ &= 0 = \begin{cases} a_{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \qquad (\text{by (646)}). \end{aligned}$$

Thus, Lemma 7.52 is proven in Case 2.

We have now proved Lemma 7.52 in both Cases 1 and 2. Since these two Cases cover all possibilities, this shows that Lemma 7.52 always holds. □

Now, we are ready to solve Exercise 3.22 again:

Second solution to Exercise 3.22. Let $g \in \{0, 1, \ldots, n\}$ be arbitrary. Then, $n - g \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$. Hence, Theorem 3.30 (applied to $n - g$ instead of n) yields

$$\binom{X+Y}{n-g} = \sum_{k=0}^{n-g} \binom{X}{k} \binom{Y}{n-g-k}$$

(an equality between two polynomials in X and Y). Substituting g and $X-g$ for X and Y in this equality, we obtain

$$\binom{g+(X-g)}{n-g} = \sum_{k=0}^{n-g} \binom{g}{k} \binom{X-g}{n-g-k} = \sum_{i=0}^{n-g} \binom{g}{i} \binom{X-g}{n-g-i}$$

(here, we have renamed the summation index k as i). Since $g+(X-g) = X$, this rewrites as

$$\binom{X}{n-g} = \sum_{i=0}^{n-g} \binom{g}{i} \binom{X-g}{n-g-i}. \qquad (647)$$

Now, let us forget that we fixed g. We thus have shown that (647) holds for every $g \in \{0, 1, \ldots, n\}$.

On the other hand, for every $k \in \mathbb{N}$ and $i \in \mathbb{N}$ satisfying $k+i \leq n$, we have

$$\binom{X}{k} \binom{X-k}{n-k-i} = \binom{X}{n-i} \binom{n-i}{k} \qquad (648)$$

[326].

[326] *Proof of (648):* Let $k \in \mathbb{N}$ and $i \in \mathbb{N}$ be such that $k+i \leq n$. From $k+i \leq n$, we obtain $n-i \geq k$. Thus, $n-i \geq k \geq 0$, so that $n-i \in \mathbb{N}$. Hence, Proposition 3.26 **(f)** (applied to $n-i$ and k instead of i and a) shows that

$$\binom{X}{n-i} \binom{n-i}{k} = \binom{X}{k} \binom{X-k}{(n-i)-k} = \binom{X}{k} \binom{X-k}{n-k-i}$$

(since $(n-i)-k = n-k-i$). This proves (648).

Now,

$$\sum_{k=0}^{n}(-1)^k\binom{X}{k}\underbrace{\binom{X}{n-k}}_{\substack{=\sum_{i=0}^{n-k}\binom{k}{i}\binom{X-k}{n-k-i}\\ \text{(by (647) (applied to } g=k))}}$$

$$=\sum_{k=0}^{n}(-1)^k\binom{X}{k}\left(\sum_{i=0}^{n-k}\binom{k}{i}\binom{X-k}{n-k-i}\right)$$

$$=\underbrace{\sum_{k=0}^{n}\sum_{i=0}^{n-k}}_{\substack{=\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ k\leq n;\ i\leq n-k}}=\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ k\leq n;\ k+i\leq n}}\\ \text{(since the condition } i\leq n-k \text{ is equivalent to } k+i\leq n)}}(-1)^k\underbrace{\binom{X}{k}\binom{k}{i}}_{=\binom{k}{i}\binom{X}{k}}\binom{X-k}{n-k-i}$$

$$=\underbrace{\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ k\leq n;\ k+i\leq n}}}_{\substack{=\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ k+i\leq n}}\\ \text{(since the condition } (k\leq n \text{ and } k+i\leq n) \text{ is equivalent to } k+i\leq n\\ \text{(because the condition } k\leq n \text{ follows from } k+i\leq n))}}(-1)^k\binom{k}{i}\underbrace{\binom{X}{k}\binom{X-k}{n-k-i}}_{\substack{=\binom{X}{n-i}\binom{n-i}{k}\\ \text{(by (648))}}}$$

$$=\underbrace{\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ k+i\leq n}}}_{\substack{=\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ i\leq n;\ k+i\leq n}}\\ \text{(since the condition } k+i\leq n \text{ is equivalent to } (i\leq n \text{ and } k+i\leq n)\\ \text{(because the condition } i\leq n \text{ follows from } k+i\leq n))}}(-1)^k\binom{k}{i}\binom{X}{n-i}\binom{n-i}{k}$$

$$= \underbrace{\sum_{\substack{(k,i)\in\mathbb{N}^2;\\ i\leq n;\ k+i\leq n}}}_{\substack{=\sum_{\substack{(i,k)\in\mathbb{N}^2;\\ i\leq n;\ k+i\leq n}} = \sum_{\substack{(i,k)\in\mathbb{N}^2;\\ i\leq n;\ k\leq n-i}} \\ \text{(since the condition } k+i\leq n \text{ is equivalent to } k\leq n-i\text{)}}} (-1)^k \binom{k}{i}\binom{X}{n-i}\binom{n-i}{k}$$

$$= \underbrace{\sum_{\substack{(i,k)\in\mathbb{N}^2;\\ i\leq n;\ k\leq n-i}}}_{=\sum_{i=0}^{n}\sum_{k=0}^{n-i}} (-1)^k \binom{k}{i}\binom{X}{n-i}\binom{n-i}{k}$$

$$= \sum_{i=0}^{n}\sum_{k=0}^{n-i} (-1)^k \binom{k}{i}\binom{X}{n-i}\binom{n-i}{k} = \sum_{i=0}^{n}\binom{X}{n-i} \underbrace{\sum_{k=0}^{n-i}(-1)^k\binom{k}{i}\binom{n-i}{k}}_{\substack{=(-1)^i[n-i=i] \\ \text{(by Lemma 7.51} \\ \text{(applied to } m=n-i\text{))}}}$$

$$= \sum_{i=0}^{n} \underbrace{\binom{X}{n-i}(-1)^i}_{=(-1)^i\binom{X}{n-i}} [n-i=i]$$

$$= \sum_{i=0}^{n} (-1)^i \binom{X}{n-i}[n-i=i] = \begin{cases} (-1)^{n/2}\dbinom{X}{n-n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

$$\left(\text{by Lemma 7.52, applied to } a_i = (-1)^i\binom{X}{n-i}\right)$$

$$= \begin{cases} (-1)^{n/2}\dbinom{X}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \qquad (\text{since } n-n/2 = n/2).$$

This solves Exercise 3.22. □

7.30.3. Addendum

Let us record a classical result which follows from Exercise 3.22:

Corollary 7.53. Let $n \in \mathbb{N}$. Then,

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k}^2 = \begin{cases} (-1)^{n/2}\dbinom{n}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

Proof of Corollary 7.53. Exercise 3.22 shows that

$$\sum_{k=0}^{n}(-1)^k\binom{X}{k}\binom{X}{n-k}=\begin{cases}(-1)^{n/2}\dbinom{X}{n/2}, & \text{if } n \text{ is even;}\\ 0, & \text{if } n \text{ is odd}\end{cases}.$$

Substituting n for X in this equality, we obtain

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{n}{n-k}=\begin{cases}(-1)^{n/2}\dbinom{n}{n/2}, & \text{if } n \text{ is even;}\\ 0, & \text{if } n \text{ is odd}\end{cases}.$$

Now,

$$\sum_{k=0}^{n}(-1)^k\underbrace{\binom{n}{k}^2}_{=\binom{n}{k}\binom{n}{k}}=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\underbrace{\binom{n}{k}}_{\substack{=\binom{n}{n-k}\\ \text{(by (232) (applied to } n \text{ and } k\\ \text{instead of } m \text{ and } n))}}$$
$$=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{n}{n-k}=\begin{cases}(-1)^{n/2}\dbinom{n}{n/2}, & \text{if } n \text{ is even;}\\ 0, & \text{if } n \text{ is odd}\end{cases}.$$

This proves Corollary 7.53. □

Corollary 7.53 is a well-known fact. Mike Spivey has found a combinatorial proof [Spivey12a] (which generalizes immediately to a proof of Lemma 7.49); the corollary also has appeared in `http://www.artofproblemsolving.com/community/c6h262752` .

7.31. Solution to Exercise 3.23

Exercise 3.23 **(a)** is a known fact. It appears, e.g., in [GrKnPa94, §5.3, (5.39)]. Combinatorial proofs appear in [Sved84] as well as in the math.stackexchange discussions `https://math.stackexchange.com/questions/72367` and `https://math.stackexchange.com/a/360780` . Complicated combinatorial proofs of Exercise 3.23 **(b)** have been given in [Spivey12b] and at `https://math.stackexchange.com/questions/80649` . We shall prove both parts of Exercise 3.23 algebraically, making use of Exercise 3.2 **(b)**. (This is how part **(a)** of this exercise is proven in [GrKnPa94, §5.3].)

Solution to Exercise 3.23. Let $k\in\{0,1,\ldots,n\}$. Thus, $n-k\in\{0,1,\ldots,n\}\subseteq\mathbb{N}$. Hence, Exercise 3.2 **(b)** (applied to $n-k$ instead of n) yields

$$\binom{-1/2}{n-k}=\left(\frac{-1}{4}\right)^{n-k}\binom{2(n-k)}{n-k}. \qquad (649)$$

But $k \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$. Hence, Exercise 3.2 **(b)** (applied to k instead of n) yields

$$\binom{-1/2}{k} = \left(\frac{-1}{4}\right)^k \binom{2k}{k}. \tag{650}$$

Multiplying the equalities (650) and (649), we obtain

$$\begin{aligned}\binom{-1/2}{k}\binom{-1/2}{n-k} &= \left(\frac{-1}{4}\right)^k \binom{2k}{k} \left(\frac{-1}{4}\right)^{n-k} \binom{2(n-k)}{n-k} \\ &= \underbrace{\left(\frac{-1}{4}\right)^k \left(\frac{-1}{4}\right)^{n-k}}_{\substack{=\left(\frac{-1}{4}\right)^{k+(n-k)} = \left(\frac{-1}{4}\right)^n \\ \text{(since } k+(n-k)=n)}} \binom{2k}{k}\binom{2(n-k)}{n-k} \\ &= \left(\frac{-1}{4}\right)^n \binom{2k}{k}\binom{2(n-k)}{n-k}. \end{aligned} \tag{651}$$

Now, forget that we fixed k. We thus have proven the equality (651) for each $k \in \{0, 1, \ldots, n\}$.

We also recall that $a^n b^n = (ab)^n$ for any $a, b \in \mathbb{Q}$. Applying this to $a = -4$ and $b = \frac{-1}{4}$, we obtain

$$(-4)^n \left(\frac{-1}{4}\right)^n = \left(\underbrace{(-4)\left(\frac{-1}{4}\right)}_{=1}\right)^n = 1^n = 1.$$

(a) Theorem 3.29 (applied to $x = -1/2$ and $y = -1/2$) yields

$$\binom{(-1/2) + (-1/2)}{n} = \sum_{k=0}^{n} \binom{-1/2}{k}\binom{-1/2}{n-k}.$$

Comparing this with

$$\begin{aligned}\binom{(-1/2) + (-1/2)}{n} &= \binom{-1}{n} && \left(\text{since } (-1/2) + (-1/2) = -1\right) \\ &= (-1)^n && (\text{by Corollary 3.17}),\end{aligned}$$

we obtain

$$(-1)^n = \sum_{k=0}^{n} \underbrace{\binom{-1/2}{k}\binom{-1/2}{n-k}}_{\substack{=\left(\frac{-1}{4}\right)^n \binom{2k}{k}\binom{2(n-k)}{n-k} \\ \text{(by (651))}}} = \sum_{k=0}^{n} \left(\frac{-1}{4}\right)^n \binom{2k}{k}\binom{2(n-k)}{n-k}$$
$$= \left(\frac{-1}{4}\right)^n \sum_{k=0}^{n} \binom{2k}{k}\binom{2(n-k)}{n-k}.$$

Multiplying both sides of this equality by $(-4)^n$, we obtain

$$(-4)^n (-1)^n = \underbrace{(-4)^n \left(\frac{-1}{4}\right)^n}_{=1} \sum_{k=0}^{n} \binom{2k}{k}\binom{2(n-k)}{n-k} = \sum_{k=0}^{n} \binom{2k}{k}\binom{2(n-k)}{n-k}.$$

Hence,

$$\sum_{k=0}^{n} \binom{2k}{k}\binom{2(n-k)}{n-k} = (-4)^n (-1)^n = \Big(\underbrace{(-4)(-1)}_{=4}\Big)^n = 4^n.$$

This solves Exercise 3.23 **(a)**.

(b) Exercise 3.22 yields

$$\sum_{k=0}^{n} (-1)^k \binom{X}{k}\binom{X}{n-k} = \begin{cases} (-1)^{n/2} \binom{X}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

(an identity between polynomials in $\mathbb{Q}[X]$). If we substitute $-1/2$ for X in this equality, we obtain

$$\sum_{k=0}^{n} (-1)^k \binom{-1/2}{k}\binom{-1/2}{n-k} = \begin{cases} (-1)^{n/2} \binom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

Hence,

$$\begin{cases} (-1)^{n/2} \binom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} = \sum_{k=0}^{n} (-1)^k \underbrace{\binom{-1/2}{k}\binom{-1/2}{n-k}}_{\substack{=\left(\frac{-1}{4}\right)^n \binom{2k}{k}\binom{2(n-k)}{n-k} \\ \text{(by (651))}}}$$
$$= \sum_{k=0}^{n} (-1)^k \left(\frac{-1}{4}\right)^n \binom{2k}{k}\binom{2(n-k)}{n-k}$$
$$= \left(\frac{-1}{4}\right)^n \sum_{k=0}^{n} (-1)^k \binom{2k}{k}\binom{2(n-k)}{n-k}.$$

Multiplying both sides of this equality by $(-4)^n$, we obtain

$$(-4)^n \begin{cases} (-1)^{n/2} \dbinom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$
$$= \underbrace{(-4)^n \left(\frac{-1}{4}\right)^n}_{=1} \sum_{k=0}^{n} (-1)^k \binom{2k}{k} \binom{2(n-k)}{n-k} = \sum_{k=0}^{n} (-1)^k \binom{2k}{k} \binom{2(n-k)}{n-k}.$$

Hence,

$$\begin{aligned} &\sum_{k=0}^{n} (-1)^k \binom{2k}{k} \binom{2(n-k)}{n-k} \\ &= (-4)^n \begin{cases} (-1)^{n/2} \dbinom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \\ &= \begin{cases} (-4)^n \cdot (-1)^{n/2} \dbinom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ (-4)^n \cdot 0, & \text{if } n \text{ is odd} \end{cases} \\ &= \begin{cases} (-4)^n \cdot (-1)^{n/2} \dbinom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} \end{aligned} \qquad (652)$$

(since $(-4)^n \cdot 0 = 0$).

Now, let us assume that n is even. Thus, $n/2 \in \mathbb{N}$. Hence, $2^n = 2^{2(n/2)} = \left(2^2\right)^{n/2} = 4^{n/2}$ (since $2^2 = 4$). From $n - n/2 = n/2$, we obtain $4^{n-n/2} = 4^{n/2} = 2^n$. But recall that $n/2 \in \mathbb{N}$. Hence, Exercise 3.2 **(b)** (applied to $n/2$ instead of n) yields

$$\binom{-1/2}{n/2} = \underbrace{\left(\frac{-1}{4}\right)^{n/2}}_{=\frac{(-1)^{n/2}}{4^{n/2}}} \underbrace{\binom{2(n/2)}{n/2}}_{\substack{=\binom{n}{n/2} \\ (\text{since } 2(n/2)=n)}} = \frac{(-1)^{n/2}}{4^{n/2}} \binom{n}{n/2}.$$

Hence,

$$\begin{aligned}
&\underbrace{(-4)^n}_{=(-1)^n\cdot 4^n} \cdot (-1)^{n/2} \underbrace{\binom{-1/2}{n/2}}_{=\frac{(-1)^{n/2}}{4^{n/2}}\binom{n}{n/2}} \\
&= (-1)^n \cdot 4^n \cdot (-1)^{n/2} \frac{(-1)^{n/2}}{4^{n/2}} \binom{n}{n/2} \\
&= \underbrace{(-1)^n}_{\substack{=1 \\ \text{(since } n \text{ is even)}}} \cdot \underbrace{(-1)^{n/2}(-1)^{n/2}}_{\substack{=(-1)^{n/2+n/2}=1 \\ \text{(since } n/2+n/2=n \text{ is even)}}} \cdot \underbrace{\frac{4^n}{4^{n/2}}}_{=4^{n-n/2}=2^n} \binom{n}{n/2} \\
&= 2^n \binom{n}{n/2}.
\end{aligned}$$

Now, forget that we assumed that n is even. We thus have proven that if n is even, then $(-4)^n \cdot (-1)^{n/2} \binom{-1/2}{n/2} = 2^n \binom{n}{n/2}$. Hence,

$$\begin{cases} (-4)^n \cdot (-1)^{n/2} \dbinom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} = \begin{cases} 2^n \dbinom{n}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

Thus, (652) becomes

$$\begin{aligned}
&\sum_{k=0}^{n} (-1)^k \binom{2k}{k} \binom{2(n-k)}{n-k} \\
&= \begin{cases} (-4)^n \cdot (-1)^{n/2} \dbinom{-1/2}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases} = \begin{cases} 2^n \dbinom{n}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.
\end{aligned}$$

This solves Exercise 3.23 **(b)**. □

7.32. Solution to Exercise 3.24

Solution to Exercise 3.24. Proposition 3.20 (applied to $2m$ and m instead of m and n) yields $\binom{2m}{m} \in \mathbb{Z}$. Similarly, we can find $\binom{2m-1}{m-1} \in \mathbb{Z}$. Thus, it makes sense to speak of $\binom{2m}{m}$ or $\binom{2m-1}{m-1}$ being even.

Proposition 3.22 (applied to $2m$ and m instead of m and n) yields $\binom{2m}{m} = \underbrace{\frac{2m}{m}}_{=2} \binom{2m-1}{m-1} = 2\binom{2m-1}{m-1}$.

(a) We have $\dbinom{2m-1}{m-1} \in \mathbb{Z}$. Thus, the integer $2\dbinom{2m-1}{m-1}$ is even. In view of $\dbinom{2m}{m} = 2\dbinom{2m-1}{m-1}$, this rewrites as follows: The integer $\dbinom{2m}{m}$ is even. This solves Exercise 3.24 **(a)**.

(b) Assume that m is odd and satisfies $m > 1$. We have $m \equiv 1 \mod 2$ (since m is odd). Also, $m-1$ is a positive integer (since $m > 1$). Hence, Exercise 3.24 **(a)** (applied to $m-1$ instead of m) shows that $\dbinom{2(m-1)}{m-1}$ is even. In other words, $\dbinom{2(m-1)}{m-1} \equiv 0 \mod 2$.

But $2m-1 = \underbrace{m}_{>1} + m - 1 > 1 + m - 1 = m$, so that $2m-1 \geq m \geq 0$ and thus $2m-1 \in \mathbb{N}$. Hence, Proposition 3.8 (applied to $2m-1$ and m instead of m and n) yields $\dbinom{2m-1}{m} = \dbinom{2m-1}{(2m-1)-m} = \dbinom{2m-1}{m-1}$ (since $(2m-1)-m = m-1$).

But Proposition 3.22 (applied to $2m-1$ and m instead of m and n) yields

$$\binom{2m-1}{m} = \frac{2m-1}{m}\binom{(2m-1)-1}{m-1} = \frac{2m-1}{m}\binom{2(m-1)}{m-1}$$

(since $(2m-1)-1 = 2(m-1)$). Multiplying both sides of this equality by m, we find

$$m\binom{2m-1}{m} = (2m-1)\underbrace{\binom{2(m-1)}{m-1}}_{\equiv 0 \mod 2} \equiv 0 \mod 2.$$

Hence,

$$0 \equiv \underbrace{m}_{\equiv 1 \mod 2}\binom{2m-1}{m} \equiv \binom{2m-1}{m} = \binom{2m-1}{m-1} \mod 2.$$

In other words, $\dbinom{2m-1}{m-1} \equiv 0 \mod 2$. In other words, the integer $\dbinom{2m-1}{m-1}$ is even. This solves Exercise 3.24 **(b)**.

(c) Assume that m is odd and satisfies $m > 1$. Exercise 3.24 **(b)** yields that the integer $\dbinom{2m-1}{m-1}$ is even. In other words, there exists an integer z such that $\dbinom{2m-1}{m-1} = 2z$. Consider this z. Now,

$$\binom{2m}{m} = 2\underbrace{\binom{2m-1}{m-1}}_{=2z} = 2 \cdot 2z = 4z \equiv 0 \mod 4$$

(since z is an integer). This solves Exercise 3.24 **(c)**. □

7.33. Solution to Exercise 3.25

Before we solve Exercise 3.25, let us prove some straightforward identities:

Lemma 7.54. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Then, $\dbinom{m+n}{m} = \dbinom{m+n}{n}$.

Proof of Lemma 7.54. We have $m+n \in \mathbb{N}$ (since $m \in \mathbb{N}$ and $n \in \mathbb{N}$). Also, $n \in \mathbb{N}$, so that $n \geq 0$. Now, $m + \underbrace{n}_{\geq 0} \geq m$. Hence, Proposition 3.8 (applied to $m+n$ and m instead of m and n) yields $\dbinom{m+n}{m} = \dbinom{m+n}{(m+n)-m} = \dbinom{m+n}{n}$ (since $(m+n)-m = n$). This proves Lemma 7.54. □

Lemma 7.55. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m,n\}$. Let $k \in \{-p, -p+1, \ldots, p\}$. Then,

$$\binom{2m}{m+k}\binom{2n}{n-k} = \frac{(2m)!\,(2n)!}{(m+n)!^2} \cdot \binom{m+n}{m+k}\binom{m+n}{n+k}.$$

Proof of Lemma 7.55. Clearly, $p = \min\{m,n\} \in \mathbb{N}$ (since $m \in \mathbb{N}$ and $n \in \mathbb{N}$). We have $k \in \{-p, -p+1, \ldots, p\}$. Thus, k is an integer satisfying $-p \leq k \leq p$.

We have $p = \min\{m,n\} \leq m$. Hence, $k \leq p \leq m$. Thus, $m - k \geq 0$. Also, $-p \leq k$, so that $k \geq -\underbrace{p}_{\leq m} \geq -m$. Hence, $k + m \geq 0$.

We have $p = \min\{m,n\} \leq n$. Hence, $k \leq p \leq n$. Thus, $n - k \geq 0$. Also, $-p \leq k$, so that $k \geq -\underbrace{p}_{\leq n} \geq -n$. Hence, $k + n \geq 0$.

From $n - k \geq 0$, we obtain $n - k \in \mathbb{N}$ (since $n - k$ is an integer). Also, $m + n \in \mathbb{N}$ (since $m \in \mathbb{N}$ and $n \in \mathbb{N}$). Furthermore, $n + k = k + n \geq 0$. Hence, $n + k \in \mathbb{N}$ (since $n + k$ is an integer). Similarly, $m + k \in \mathbb{N}$.

Now, $m + n \geq m + k$ (since $m + \underbrace{k}_{\leq n} \leq m + n$). Hence, Proposition 3.4 (applied to $m + n$ and $m + k$ instead of m and n) yields

$$\binom{m+n}{m+k} = \frac{(m+n)!}{(m+k)!\,((m+n)-(m+k))!} = \frac{(m+n)!}{(m+k)!\,(n-k)!} \tag{653}$$

(since $(m+n)-(m+k) = n-k$).

Also, $m + n \geq n + k$ (since $n + \underbrace{k}_{\leq m} \leq n + m = m + n$). Hence, Proposition 3.4 (applied to $m + n$ and $n + k$ instead of m and n) yields

$$\binom{m+n}{n+k} = \frac{(m+n)!}{(n+k)!\,((m+n)-(n+k))!} = \frac{(m+n)!}{(n+k)!\,(m-k)!} \tag{654}$$

(since $(m+n)-(n+k)=m-k$).

Now,

$$\begin{aligned}
&\frac{(2m)!\,(2n)!}{(m+n)!^2}\cdot \underbrace{\binom{m+n}{m+k}}_{\substack{=\frac{(m+n)!}{(m+k)!\,(n-k)!}\\ \text{(by (653))}}} \underbrace{\binom{m+n}{n+k}}_{\substack{=\frac{(m+n)!}{(n+k)!\,(m-k)!}\\ \text{(by (654))}}}\\
&=\frac{(2m)!\,(2n)!}{(m+n)!^2}\cdot\frac{(m+n)!}{(m+k)!\,(n-k)!}\cdot\frac{(m+n)!}{(n+k)!\,(m-k)!}\\
&=\frac{(2m)!\,(2n)!}{(m+k)!\,(n-k)!\,(n+k)!\,(m-k)!}.
\end{aligned} \tag{655}$$

Also, $2m\in\mathbb{N}$ (since $m\in\mathbb{N}$) and $2m\geq m+k$ (since $2m-(m+k)=m-k\geq 0$). Hence, Proposition 3.4 (applied to $2m$ and $m+k$ instead of m and n) yields

$$\binom{2m}{m+k}=\frac{(2m)!}{(m+k)!\,(2m-(m+k))!}=\frac{(2m)!}{(m+k)!\,(m-k)!} \tag{656}$$

(since $2m-(m+k)=m-k$).

Moreover, $2n\in\mathbb{N}$ (since $n\in\mathbb{N}$) and $2n\geq n-k$ (since $2n-(n-k)=k+n\geq 0$). Hence, Proposition 3.4 (applied to $2n$ and $n-k$ instead of m and n) yields

$$\binom{2n}{n-k}=\frac{(2n)!}{(n-k)!\,(2n-(n-k))!}=\frac{(2n)!}{(n-k)!\,(n+k)!} \tag{657}$$

(since $2n-(n-k)=n+k$).

Multiplying the equalities (656) and (657), we obtain

$$\begin{aligned}
\binom{2m}{m+k}\binom{2n}{n-k}&=\frac{(2m)!}{(m+k)!\,(m-k)!}\cdot\frac{(2n)!}{(n-k)!\,(n+k)!}\\
&=\frac{(2m)!\,(2n)!}{(m+k)!\,(n-k)!\,(n+k)!\,(m-k)!}\\
&=\frac{(2m)!\,(2n)!}{(m+n)!^2}\cdot\binom{m+n}{m+k}\binom{m+n}{n+k}
\end{aligned}$$

(by (655)). This proves Lemma 7.55. □

Solution to Exercise 3.25. Let us begin with parts **(e)**, **(f)**, **(g)**, **(h)** and **(i)** of Exercise 3.25.

(e) Let $m\in\mathbb{N}$. Thus, $2m\geq m\geq 0$. Hence, Proposition 3.4 (applied to $2m$ and m instead of m and n) yields $\binom{2m}{m}=\frac{(2m)!}{m!\,(2m-m)!}=\frac{(2m)!}{m!m!}$. But the definition of

$T(m,0)$ yields

$$T(m,0) = \frac{(2m)!\,(2\cdot 0)!}{m!0!\,(m+0)!} = \frac{(2m)!\cdot 1}{m!\cdot 1\cdot (m+0)!} \qquad \left(\text{since } (2\cdot 0)! = 0! = 1 \text{ and } 0! = 1\right)$$
$$= \frac{(2m)!}{m!\,(m+0)!} = \frac{(2m)!}{m!m!} = \binom{2m}{m}.$$

This solves Exercise 3.25 **(e)**.

(f) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

From $m \in \mathbb{N}$, we obtain $2m \geq m$. Hence, Proposition 3.4 (applied to $2m$ and m instead of m and n) yields $\binom{2m}{m} = \frac{(2m)!}{m!\,(2m-m)!} = \frac{(2m)!}{m!m!}$. Similarly, $\binom{2n}{n} = \frac{(2n)!}{n!n!}$. Also, $m + \underbrace{n}_{\geq 0} \geq m$. Hence, Proposition 3.4 (applied to $m+n$ and m instead of m and n) yields $\binom{m+n}{m} = \frac{(m+n)!}{m!\,((m+n)-m)!} = \frac{(m+n)!}{m!n!}$. Thus, the rational number $\binom{m+n}{m}$ is nonzero (since $\frac{(m+n)!}{m!n!}$ is clearly nonzero). Hence, the fraction $\frac{\binom{2m}{m}\binom{2n}{n}}{\binom{m+n}{m}}$ is well-defined.

Now,

$$\frac{\binom{2m}{m}\binom{2n}{n}}{\binom{m+n}{m}} = \frac{\frac{(2m)!}{m!m!}\cdot\frac{(2n)!}{n!n!}}{\left(\frac{(m+n)!}{m!n!}\right)} \qquad \left(\begin{array}{c}\text{since } \binom{2m}{m} = \frac{(2m)!}{m!m!} \text{ and } \binom{2n}{n} = \frac{(2n)!}{n!n!} \\ \text{and } \binom{m+n}{m} = \frac{(m+n)!}{m!n!}\end{array}\right)$$
$$= \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}. \tag{658}$$

But the definition of $T(m,n)$ yields

$$T(m,n) = \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}.$$

Comparing this with (658), we obtain $T(m,n) = \frac{\binom{2m}{m}\binom{2n}{n}}{\binom{m+n}{m}}$. This solves Exercise 3.25 **(f)**.

(g) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. The definition of $T(n,m)$ yields

$$T(n,m) = \frac{(2n)!\,(2m)!}{n!m!\,(n+m)!} = \frac{(2m)!\,(2n)!}{m!n!\,(n+m)!} = \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!} \tag{659}$$

(since $n+m=m+n$). But the definition of $T(m,n)$ yields

$$T(m,n)=\frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}.$$

Comparing this with (659), we obtain $T(m,n)=T(n,m)$. This solves Exercise 3.25 **(g)**.

(h) Let us forget that we fixed m, n and p. We shall first prove some auxiliary observations:

Observation 1: Let $m\in\mathbb{N}$ and $n\in\mathbb{N}$ be such that $m\geq n$. Then,

$$\sum_{k=-n}^{n}(-1)^k\binom{m+n}{m+k}\binom{m+n}{n+k}=\binom{m+n}{m}.$$

[*Proof of Observation 1:* Every $k\in\{-n,-n+1,\ldots,n\}$ satisfies

$$\binom{m+n}{n-k}=\binom{m+n}{m+k} \qquad (660)$$

[327].

Exercise 3.22 (applied to $2n$ instead of n) yields

$$\begin{aligned}\sum_{k=0}^{2n}(-1)^k\binom{X}{k}\binom{X}{2n-k}&=\begin{cases}(-1)^{2n/2}\dbinom{X}{2n/2}, & \text{if } 2n \text{ is even;}\\ 0, & \text{if } 2n \text{ is odd}\end{cases}\\&=(-1)^{2n/2}\binom{X}{2n/2}\qquad(\text{since } 2n \text{ is even})\\&=(-1)^n\binom{X}{n}\qquad(\text{since } 2n/2=n)\end{aligned}$$

(an identity between polynomials in $\mathbb{Q}[X]$). Substituting $m+n$ for X in this equality, we obtain

$$\sum_{k=0}^{2n}(-1)^k\binom{m+n}{k}\binom{m+n}{2n-k}=(-1)^n\binom{m+n}{n}.$$

[327] *Proof of (660):* Let $k\in\{-n,-n+1,\ldots,n\}$. Thus, k is an integer satisfying $-n\leq k\leq n$. From $k\leq n$, we obtain $n-k\geq 0$. Hence, $n-k\in\mathbb{N}$ (since $n-k$ is an integer). Also, $(m+n)-(n-k)=\underbrace{m}_{\geq n}+\underbrace{k}_{\substack{\geq -n\\(\text{since } -n\leq k)}}\geq n+(-n)=0$; in other words, $m+n\geq n-k$. Hence, Proposition 3.8 (applied to $m+n$ and $n-k$ instead of m and n) yields $\binom{m+n}{n-k}=\binom{m+n}{(m+n)-(n-k)}=\binom{m+n}{m+k}$. This proves (660).

Hence,

$$(-1)^n \binom{m+n}{n} = \sum_{k=0}^{2n} (-1)^k \binom{m+n}{k} \binom{m+n}{2n-k}$$
$$= \sum_{k=-n}^{n} \underbrace{(-1)^{k+n}}_{=(-1)^k(-1)^n} \underbrace{\binom{m+n}{k+n}}_{=\binom{m+n}{n+k}} \underbrace{\binom{m+n}{2n-(k+n)}}_{\substack{=\binom{m+n}{n-k} \\ (\text{since } 2n-(k+n)=n-k)}}$$
(here, we have substituted $k+n$ for k in the sum)
$$= (-1)^n \sum_{k=-n}^{n} (-1)^k \binom{m+n}{n+k} \binom{m+n}{n-k}.$$

Dividing both sides of this equality by $(-1)^n$, we obtain

$$\binom{m+n}{n} = \sum_{k=-n}^{n} (-1)^k \binom{m+n}{n+k} \underbrace{\binom{m+n}{n-k}}_{\substack{=\binom{m+n}{m+k} \\ (\text{by (660)})}} = \sum_{k=-n}^{n} (-1)^k \underbrace{\binom{m+n}{n+k}\binom{m+n}{m+k}}_{=\binom{m+n}{m+k}\binom{m+n}{n+k}}$$
$$= \sum_{k=-n}^{n} (-1)^k \binom{m+n}{m+k} \binom{m+n}{n+k}.$$

In other words,

$$\sum_{k=-n}^{n} (-1)^k \binom{m+n}{m+k} \binom{m+n}{n+k} = \binom{m+n}{n} = \binom{m+n}{m}$$

(by Lemma 7.54). This proves Observation 1.]

Observation 2: Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$ be such that $m \leq n$. Then,

$$\sum_{k=-m}^{m} (-1)^k \binom{m+n}{m+k} \binom{m+n}{n+k} = \binom{m+n}{m}.$$

[*Proof of Observation 2:* We have $m \leq n$, and thus $n \geq m$. Hence, Observation 1 (applied to n and m instead of m and n) yields

$$\sum_{k=-m}^{m} (-1)^k \binom{n+m}{n+k} \binom{n+m}{m+k} = \binom{n+m}{n}.$$

Comparing this with

$$\sum_{k=-m}^{m} (-1)^k \underbrace{\binom{n+m}{n+k}\binom{n+m}{m+k}}_{=\binom{n+m}{m+k}\binom{n+m}{n+k}} = \sum_{k=-m}^{m} (-1)^k \binom{n+m}{m+k}\binom{n+m}{n+k},$$

we obtain

$$\sum_{k=-m}^{m} (-1)^k \binom{n+m}{m+k}\binom{n+m}{n+k} = \binom{n+m}{n}.$$

In view of $n+m = m+n$, this rewrites as

$$\sum_{k=-m}^{m} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \binom{m+n}{n}.$$

Comparing this with

$$\binom{m+n}{m} = \binom{m+n}{n} \qquad \text{(by Lemma 7.54)},$$

we obtain

$$\sum_{k=-m}^{m} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \binom{m+n}{m}.$$

This proves Observation 2.]

Let us now come back to the solution of Exercise 3.25 **(h)**. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m, n\}$. We are in one of the following two cases:

Case 1: We have $m \geq n$.

Case 2: We have $m < n$.

Let us first consider Case 1. In this case, we have $m \geq n$. Now, $p = \min\{m, n\} = n$ (since $m \geq n$). Hence,

$$\sum_{k=-p}^{p} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \sum_{k=-n}^{n} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \binom{m+n}{m}$$

(by Observation 1). Hence, Exercise 3.25 **(h)** is proven in Case 1.

Let us first consider Case 2. In this case, we have $m < n$. Hence, $m \leq n$. Now, $p = \min\{m, n\} = m$ (since $m \leq n$). Hence,

$$\sum_{k=-p}^{p} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \sum_{k=-m}^{m} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \binom{m+n}{m}$$

(by Observation 2). Hence, Exercise 3.25 **(h)** is proven in Case 2.

We have thus proven Exercise 3.25 **(h)** in each of the two Cases 1 and 2. Thus, Exercise 3.25 **(h)** is solved in all cases.

(i) We have $m+n \geq m$ (since $n \geq 0$). Hence, Proposition 3.4 (applied to $m+n$ and m instead of m and n) yields $\dbinom{m+n}{m} = \dfrac{(m+n)!}{m!\left((m+n)-m\right)!} = \dfrac{(m+n)!}{m!n!}$.

But Exercise 3.25 **(h)** yields

$$\sum_{k=-p}^{p} (-1)^k \binom{m+n}{m+k} \binom{m+n}{n+k} = \binom{m+n}{m} = \frac{(m+n)!}{m!n!}. \tag{661}$$

Now,

$$\begin{aligned}
\sum_{k=-p}^{p} (-1)^k & \underbrace{\binom{2m}{m+k}\binom{2n}{n-k}}_{\substack{=\frac{(2m)!\,(2n)!}{(m+n)!^2}\cdot\binom{m+n}{m+k}\binom{m+n}{n+k} \\ \text{(by Lemma 7.55)}}} \\
&= \sum_{k=-p}^{p} (-1)^k \frac{(2m)!\,(2n)!}{(m+n)!^2} \cdot \binom{m+n}{m+k}\binom{m+n}{n+k} \\
&= \frac{(2m)!\,(2n)!}{(m+n)!^2} \cdot \underbrace{\sum_{k=-p}^{p} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k}}_{\substack{=\frac{(m+n)!}{m!n!} \\ \text{(by (661))}}} \\
&= \frac{(2m)!\,(2n)!}{(m+n)!^2} \cdot \frac{(m+n)!}{m!n!} = \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}.
\end{aligned}$$

Comparing this with

$$T(m,n) = \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!} \qquad \text{(by the definition of } T(m,n)\text{)},$$

we obtain $T(m,n) = \sum_{k=-p}^{p} (-1)^k \dbinom{2m}{m+k}\dbinom{2n}{n-k}$. This solves Exercise 3.25 **(i)**.

(a) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Then, $2(m+1) = 2m+2$, so that

$$\begin{aligned}
(2(m+1))! = (2m+2)! &= 1 \cdot 2 \cdot \cdots \cdot (2m+2) \\
&= \underbrace{(1 \cdot 2 \cdot \cdots \cdot (2m))}_{=(2m)!} \cdot (2m+1) \cdot \underbrace{(2m+2)}_{=2(m+1)} \\
&= (2m)! \cdot (2m+1) \cdot 2(m+1).
\end{aligned} \tag{662}$$

Now, the definition of $T(m+1,n)$ yields

$$\begin{aligned}
T(m+1,n) &= \frac{(2(m+1))!\,(2n)!}{(m+1)!n!\,(m+1+n)!} = \frac{(2m)!\cdot(2m+1)\cdot 2(m+1)\cdot(2n)!}{m!\cdot(m+1)\cdot n!\cdot(m+n)!\cdot(m+n+1)} \\
&\qquad \left(\begin{array}{c}\text{by (662) and because of } (m+1)! = m!\cdot(m+1) \\ \text{and } (m+1+n)! = (m+n+1)! = (m+n)!\cdot(m+n+1)\end{array}\right) \\
&= \frac{2\cdot(2m+1)}{m+n+1}\cdot \underbrace{\frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}}_{\substack{=T(m,n)\\ \text{(by the definition of } T(m,n))}} \\
&= \frac{2\cdot(2m+1)}{m+n+1}\cdot T(m,n). \qquad (663)
\end{aligned}$$

The same argument (but with m and n replaced by n and m) shows that

$$T(n+1,m) = \frac{2\cdot(2n+1)}{n+m+1}\cdot T(n,m).$$

But Exercise 3.25 **(g)** yields $T(m,n) = T(n,m)$. Also, Exercise 3.25 **(g)** (applied to $n+1$ instead of n) yields

$$T(m,n+1) = T(n+1,m) = \underbrace{\frac{2\cdot(2n+1)}{n+m+1}}_{=\frac{2\cdot(2n+1)}{m+n+1}}\cdot \underbrace{T(n,m)}_{=T(m,n)} = \frac{2\cdot(2n+1)}{m+n+1}\cdot T(m,n).$$

Adding this equality to (663), we obtain

$$\begin{aligned}
T(m+1,n)+T(m,n+1) &= \frac{2\cdot(2m+1)}{m+n+1}\cdot T(m,n) + \frac{2\cdot(2n+1)}{m+n+1}\cdot T(m,n) \\
&= \underbrace{\left(\frac{2\cdot(2m+1)}{m+n+1}+\frac{2\cdot(2n+1)}{m+n+1}\right)}_{\substack{=4\\ \text{(by straightforward computation)}}}\cdot T(m,n) = 4T(m,n).
\end{aligned}$$

In other words, $4T(m,n) = T(m+1,n) + T(m,n+1)$. This solves Exercise 3.25 **(a)**.

Before we come to the solution of Exercise 3.25 **(b)**, let us observe something trivial:

> *Observation 3:* Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Then, $T(m,n) > 0$.

[*Proof of Observation 3:* This follows immediately from the definition of $T(m,n)$ (since the numbers $(2m)!$, $(2n)!$, $m!$, $n!$ and $(m+n)!$ are positive).]

(b) *First solution to Exercise 3.25 **(b)**:* Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m, n\}$. Exercise 3.25 **(i)** yields

$$T(m,n) = \sum_{k=-p}^{p} (-1)^k \binom{2m}{m+k} \binom{2n}{n-k}.$$

The right-hand side of this equality is clearly an integer (since the binomial coefficients appearing in it are integers[328]). Thus, so is the left-hand side. In other words, $T(m,n)$ is an integer. But Observation 3 yields $T(m,n) > 0$. Hence, $T(m,n)$ is a positive integer (since $T(m,n)$ is an integer). Therefore, $T(m,n) \in \mathbb{N}$. This solves Exercise 3.25 **(b)**.

*Second solution to Exercise 3.25 **(b)**:* We shall prove Exercise 3.25 **(b)** by induction on n:

Induction base: We have $T(m,0) \in \mathbb{N}$ for every $m \in \mathbb{N}$ [329]. In other words, Exercise 3.25 **(b)** holds for $n = 0$. This completes the induction base.

Induction step: Let $N \in \mathbb{N}$. Assume that Exercise 3.25 **(b)** holds for $n = N$. We must prove that Exercise 3.25 **(b)** holds for $n = N + 1$.

We have assumed that Exercise 3.25 **(b)** holds for $n = N$. In other words, we have

$$T(m,N) \in \mathbb{N} \qquad \text{for every } m \in \mathbb{N}. \tag{664}$$

Now, let $m \in \mathbb{N}$. We shall show that $T(m, N+1) \in \mathbb{N}$.

Indeed, (664) yields $T(m,N) \in \mathbb{N} \subseteq \mathbb{Z}$. But (664) (applied to $m+1$ instead of m) yields $T(m+1,N) \in \mathbb{N} \subseteq \mathbb{Z}$. Hence, $4T(m,N) - T(m+1,N) \in \mathbb{Z}$ (since both $T(m,N)$ and $T(m+1,N)$ belong to $\mathbb{Z}$).

Exercise 3.25 **(a)** (applied to $n = N$) yields $4T(m,N) = T(m+1,N) + T(m,N+1)$. Hence, $T(m,N+1) = 4T(m,N) - T(m+1,N) \in \mathbb{Z}$.

But Observation 3 (applied to $n = N+1$) yields $T(m,N+1) > 0$. Combining this with $T(m,N+1) \in \mathbb{Z}$, we conclude that $T(m,N+1) \in \mathbb{N}$.

Now, forget that we fixed m. We thus have proven that $T(m,N+1) \in \mathbb{N}$ for every $m \in \mathbb{N}$. In other words, Exercise 3.25 **(b)** holds for $n = N+1$. This completes the induction step. Thus, the induction proof of Exercise 3.25 **(b)** is complete.

(c) For every $m \in \mathbb{N}$ and $n \in \mathbb{N}$, we have $T(m,n) \in \mathbb{N}$ (by Exercise 3.25 **(b)**). Thus, for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$, the number $T(m,n)$ is an integer. Hence, speaking of "the integer $T(m,n)$" in Exercise 3.25 **(c)** makes sense.

We shall prove Exercise 3.25 **(c)** by induction on n:

Induction base: If $m \in \mathbb{N}$ is such that $(m,0) \neq (0,0)$, then the integer $T(m,0)$ is even[330]. In other words, Exercise 3.25 **(c)** holds for $n = 0$. This completes the induction base.

[328] according to Proposition 3.20

[329] *Proof.* Let $m \in \mathbb{N}$. Then, Exercise 3.25 **(e)** shows that $T(m,0) = \binom{2m}{m} \in \mathbb{N}$ (this follows from an application of Lemma 3.19). Qed.

[330] *Proof.* Let $m \in \mathbb{N}$ be such that $(m,0) \neq (0,0)$. Then, Exercise 3.25 **(e)** shows that $T(m,0) = \binom{2m}{m}$.

Induction step: Let $N \in \mathbb{N}$. Assume that Exercise 3.25 **(c)** holds for $n = N$. We must prove that Exercise 3.25 **(c)** holds for $n = N + 1$.

We have assumed that Exercise 3.25 **(c)** holds for $n = N$. In other words, if $m \in \mathbb{N}$ is such that $(m, N) \neq (0, 0)$, then

$$\text{the integer } T(m, N) \text{ is even.} \tag{665}$$

Now, let $m \in \mathbb{N}$ be such that $(m, N+1) \neq (0, 0)$. We shall show that the integer $T(m, N+1)$ is even.

We have $(m+1, N) \neq (0, 0)$ (since $m + 1 \neq 0$). Hence, (665) (applied to $m + 1$ instead of m) shows that the integer $T(m+1, N)$ is even. In other words, $T(m+1, N) \equiv 0 \mod 2$.

But Exercise 3.25 **(b)** (applied to $n = N$) yields $T(m, N) \in \mathbb{N}$. Hence, $4T(m, N) \equiv 0 \mod 2$.

Exercise 3.25 **(a)** (applied to $n = N$) yields $4T(m, N) = T(m+1, N) + T(m, N+1)$. Hence,

$$T(m, N+1) = \underbrace{4T(m, N)}_{\equiv 0 \mod 2} - \underbrace{T(m+1, N)}_{\equiv 0 \mod 2} \equiv 0 - 0 = 0 \mod 2.$$

In other words, the integer $T(m, N+1)$ is even.

Now, forget that we fixed m. We thus have proven that if $m \in \mathbb{N}$ is such that $(m, N+1) \neq (0, 0)$, then the integer $T(m, N+1)$ is even. In other words, Exercise 3.25 **(c)** holds for $n = N + 1$. This completes the induction step. Thus, the induction proof of Exercise 3.25 **(c)** is complete.

(d) For every $m \in \mathbb{N}$ and $n \in \mathbb{N}$, we have $T(m, n) \in \mathbb{N}$ (by Exercise 3.25 **(b)**). Thus, for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$, the number $T(m, n)$ is an integer. Hence, the divisibility statement "$4 \mid T(m, n)$" in Exercise 3.25 **(d)** makes sense.

We shall prove Exercise 3.25 **(d)** by induction on n:

Induction base: If $m \in \mathbb{N}$ is such that $m + 0$ is odd and $m + 0 > 1$, then $4 \mid T(m, 0)$ [331]. In other words, Exercise 3.25 **(d)** holds for $n = 0$. This completes the induction base.

If we had $m = 0$, then we would have $\left(\underbrace{m}_{=0}, 0\right) = (0, 0)$, which would contradict $(m, 0) \neq (0, 0)$. Hence, we cannot have $m = 0$. Thus, we must have $m \neq 0$. Therefore, m is a positive integer (since $m \in \mathbb{N}$). Hence, Exercise 3.24 **(a)** shows that the binomial coefficient $\dbinom{2m}{m}$ is even. In other words, the integer $\dbinom{2m}{m}$ is even. In view of $T(m, 0) = \dbinom{2m}{m}$, this rewrites as follows: The integer $T(m, 0)$ is even. Qed.

[331] *Proof.* Let $m \in \mathbb{N}$ be such that $m + 0$ is odd and $m + 0 > 1$. Then, Exercise 3.25 **(e)** shows that $T(m, 0) = \dbinom{2m}{m}$.

We know that $m + 0$ is odd. In other words, m is odd (since $m = m + 0$). Furthermore, m is a positive integer (since $m = m + 0 > 1 > 0$). Therefore, Exercise 3.24 **(c)** shows that $\dbinom{2m}{m} \equiv 0 \mod 4$. Thus, $T(m, 0) = \dbinom{2m}{m} \equiv 0 \mod 4$. In other words, $4 \mid T(m, 0)$. Qed.

Induction step: Let $N \in \mathbb{N}$. Assume that Exercise 3.25 **(d)** holds for $n = N$. We must prove that Exercise 3.25 **(d)** holds for $n = N + 1$.

We have assumed that Exercise 3.25 **(d)** holds for $n = N$. In other words, if $m \in \mathbb{N}$ is such that $m + N$ is odd and $m + N > 1$, then

$$4 \mid T(m, N). \tag{666}$$

Now, let $m \in \mathbb{N}$ be such that $m + (N+1)$ is odd and $m + (N+1) > 1$. We shall show that $4 \mid T(m, N+1)$.

We know that $m + (N+1)$ is odd. In other words, $(m+1) + N$ is odd (since $(m+1) + N = m + (N+1)$). Also, $(m+1) + N = m + (N+1) > 1$. Hence, (666) (applied to $m+1$ instead of m) shows that $4 \mid T(m+1, N)$. In other words, $T(m+1, N) \equiv 0 \mod 4$.

But Exercise 3.25 **(b)** (applied to $n = N$) yields $T(m, N) \in \mathbb{N}$. Hence, $4T(m, N) \equiv 0 \mod 4$.

Exercise 3.25 **(a)** (applied to $n = N$) yields $4T(m, N) = T(m+1, N) + T(m, N+1)$. Hence,

$$T(m, N+1) = \underbrace{4T(m, N)}_{\equiv 0 \mod 4} - \underbrace{T(m+1, N)}_{\equiv 0 \mod 4} \equiv 0 - 0 = 0 \mod 4.$$

In other words, $4 \mid T(m, N+1)$.

Now, forget that we fixed m. We thus have proven that if $m \in \mathbb{N}$ is such that $m + (N+1)$ is odd and $m + (N+1) > 1$, then $4 \mid T(m, N+1)$. In other words, Exercise 3.25 **(d)** holds for $n = N + 1$. This completes the induction step. Thus, the induction proof of Exercise 3.25 **(d)** is complete. □

7.34. Solution to Exercise 3.26

7.34.1. First solution

Before we solve Exercise 3.26, let us prove two basic facts in modular arithmetic:

Proposition 7.56. Let b and c be two integers such that $c > 0$. Then, there exists an $s \in \mathbb{Z}$ such that $b^{c-1} \equiv sb^c \mod c$.

Proof of Proposition 7.56. For every integer m, let $m\%c$ denote the remainder obtained when m is divided by c. Thus, every integer m satisfies

$$m\%c \in \{0, 1, \ldots, c-1\} \tag{667}$$

and

$$m\%c \equiv m \mod c. \tag{668}$$

Indeed, these two relations follow from Corollary 2.155 **(a)** (applied to $N = c$ and $n = m$).

We shall now show that two of the $c + 1$ integers $b^0\%c, b^1\%c, \ldots, b^c\%c$ are equal.

Indeed, assume the contrary (for the sake of contradiction). Thus, the $c+1$ integers $b^0\%c, b^1\%c, \ldots, b^c\%c$ are pairwise distinct. Hence,

$$\left|\left\{b^0\%c, b^1\%c, \ldots, b^c\%c\right\}\right| = c+1.$$

But we have $b^i\%c \in \{0,1,\ldots,c-1\}$ for each $i \in \{0,1,\ldots,c\}$ (by (667), applied to $m = b^i$). Thus,

$$\left\{b^0\%c, b^1\%c, \ldots, b^c\%c\right\} \subseteq \{0,1,\ldots,c-1\}.$$

Therefore,

$$\left|\left\{b^0\%c, b^1\%c, \ldots, b^c\%c\right\}\right| \leq |\{0,1,\ldots,c-1\}| = c,$$

so that

$$c \geq \left|\left\{b^0\%c, b^1\%c, \ldots, b^c\%c\right\}\right| = c+1.$$

This contradicts $c < c+1$. This contradiction shows that our assumption was wrong. Hence, we have proven that two of the $c+1$ integers $b^0\%c, b^1\%c, \ldots, b^c\%c$ are equal. In other words, there exist two distinct elements u and v of $\{0,1,\ldots,c\}$ such that $b^u\%c = b^v\%c$. Consider these u and v.

We can WLOG assume that $u \leq v$ (since otherwise, we can simply switch u with v, and nothing changes). Assume this. Thus, $u \leq v$, so that $u < v$ (since u and v are distinct).

Now, (668) (applied to $m = b^u$) yields $b^u\%c \equiv b^u \mod c$. Also, (668) (applied to $m = b^v$) yields $b^v\%c \equiv b^v \mod c$. Thus, $b^u \equiv b^u\%c = b^v\%c \equiv b^v \mod c$.

Also, $u \in \{0,1,\ldots,c\}$, so that $0 \leq u$. Also, $v \in \{0,1,\ldots,c\}$, so that $v \leq c$. Hence, $c - v \in \mathbb{N}$. Thus, b^{c-v} is a well-defined integer.

Also, $0 \leq u < v$, so that $0 \leq v-1$ (since 0 and v are integers). Thus, $v-1 \in \mathbb{N}$, so that b^{v-1} is a well-defined integer.

But $u < v$, and thus $u \leq v-1$ (since u and v are integers). Hence, $(v-1)-u \in \mathbb{N}$. Thus, $b^{(v-1)-u}$ is a well-defined integer. Set $t = b^{(v-1)-u}$. Thus, t is an integer.

Now, $v-1 = ((v-1)-u)+u$, and thus

$$\begin{aligned} b^{v-1} &= b^{((v-1)-u)+u} = \underbrace{b^{(v-1)-u}}_{=t} b^u \qquad (\text{since } 0 \leq u \leq v-1) \\ &= t \underbrace{b^u}_{\equiv b^v \bmod c} \equiv tb^v \mod c. \end{aligned}$$

Now, $c-1 = (v-1)+(c-v)$, so that

$$\begin{aligned} b^{c-1} &= b^{(v-1)+(c-v)} = \underbrace{b^{v-1}}_{\equiv tb^v \bmod c} b^{c-v} \qquad (\text{since } v-1 \in \mathbb{N} \text{ and } c-v \in \mathbb{N}) \\ &\equiv t \underbrace{b^v b^{c-v}}_{=b^{v+(c-v)}=b^c} = tb^c \mod c. \end{aligned}$$

Hence, there exists an $s \in \mathbb{Z}$ such that $b^{c-1} \equiv sb^c \mod c$ (namely, $s = t$). This proves Proposition 7.56. $\square$

Lemma 7.57. Let $n \in \mathbb{N}$. Let c be a positive integer. Let $u_0, u_1, \ldots, u_{n-1}$ be n integers. Let $v_0, v_1, \ldots, v_{n-1}$ be n integers. Let d be an integer. Assume that

$$du_i \equiv dv_i \mod c \qquad \text{for each } i \in \{0, 1, \ldots, n-1\}. \tag{669}$$

Then,

$$d \prod_{i=0}^{n-1} u_i \equiv d \prod_{i=0}^{n-1} v_i \mod c.$$

Proof of Lemma 7.57. We claim that

$$d \prod_{i=0}^{k-1} u_i \equiv d \prod_{i=0}^{k-1} v_i \mod c \tag{670}$$

for each $k \in \{0, 1, \ldots, n\}$.

[*Proof of (670):* We shall prove (670) by induction over k:

Induction base: We have $d \underbrace{\prod_{i=0}^{0-1} u_i}_{=(\text{empty product})=1} = d$. Comparing this with $d \underbrace{\prod_{i=0}^{0-1} v_i}_{=(\text{empty product})=1} = d$, we obtain $d \prod_{i=0}^{0-1} u_i = d \prod_{i=0}^{0-1} v_i$. Hence, $d \prod_{i=0}^{0-1} u_i \equiv d \prod_{i=0}^{0-1} v_i \mod c$. In other words, (670) holds for $k = 0$. This completes the induction base.

Induction step: Let $K \in \{0, 1, \ldots, n\}$ be positive. Assume that (670) holds for $k = K - 1$. We must show that (670) holds for $k = K$.

We have $K \in \{1, 2, \ldots, n\}$ (since $K \in \{0, 1, \ldots, n\}$ and K is positive). Hence, $K - 1 \in \{0, 1, \ldots, n-1\}$.

We have assumed that (670) holds for $k = K - 1$. In other words, we have

$$d \prod_{i=0}^{(K-1)-1} u_i \equiv d \prod_{i=0}^{(K-1)-1} v_i \mod c. \tag{671}$$

Now,

$$\begin{aligned} d \underbrace{\prod_{i=0}^{K-1} u_i}_{= \left(\prod_{i=0}^{(K-1)-1} u_i\right) u_{K-1}} &= \underbrace{d \left(\prod_{i=0}^{(K-1)-1} u_i\right)}_{\substack{\equiv d \prod_{i=0}^{(K-1)-1} v_i \bmod c \\ \text{(by (671))}}} u_{K-1} \\ &\equiv \left(d \prod_{i=0}^{(K-1)-1} v_i\right) u_{K-1} = \underbrace{d u_{K-1}}_{\substack{\equiv d v_{K-1} \bmod c \\ \text{(by (669), applied to } i = K-1)}} \prod_{i=0}^{(K-1)-1} v_i \\ &\equiv d v_{K-1} \prod_{i=0}^{(K-1)-1} v_i = d \underbrace{\left(\prod_{i=0}^{(K-1)-1} v_i\right) v_{K-1}}_{= \prod_{i=0}^{K-1} v_i} = d \prod_{i=0}^{K-1} v_i \bmod c. \end{aligned}$$

In other words, (670) holds for $k = K$. This completes the induction step. Thus, (670) is proven by induction.]

Now, (670) (applied to $k = n$) yields

$$d \prod_{i=0}^{n-1} u_i \equiv d \prod_{i=0}^{n-1} v_i \bmod c.$$

This proves Lemma 7.57. □

First solution to Exercise 3.26. Set $c = n!$. Notice that c is a positive integer (since $c = n! = 1 \cdot 2 \cdot \cdots \cdot n$); hence, $c - 1 \in \mathbb{N}$. Thus, $n + (c-1) \in \mathbb{N}$ (since $n \in \mathbb{N}$).

Proposition 7.56 shows that there exists an $s \in \mathbb{Z}$ such that $b^{c-1} \equiv sb^c \bmod c$. Consider this s.

For every $i \in \mathbb{Z}$, we have

$$\begin{aligned} b^{c-1}(a - bi) &= \underbrace{b^{c-1}}_{\equiv sb^c \bmod c} a - \underbrace{b^{c-1} b}_{= b^c} i \\ &\equiv sb^c a - b^c i = \underbrace{b^c}_{= b^{c-1} b} (sa - i) = b^{c-1} b (sa - i) \bmod c. \end{aligned} \tag{672}$$

Thus, in particular, (672) holds for every $i \in \{0, 1, \ldots, n-1\}$. Hence, Lemma 7.57 (applied to $d = b^{c-1}$, $u_i = a - bi$ and $v_i = b(sa - i)$) yields

$$b^{c-1} \prod_{i=0}^{n-1} (a - bi) \equiv b^{c-1} \prod_{i=0}^{n-1} (b(sa - i)) \bmod c. \tag{673}$$

For each $m \in \mathbb{Q}$, we have

$$\dbinom{m}{n} = \frac{m(m-1)\cdots(m-n+1)}{n!} = \underbrace{\frac{1}{n!}}_{\substack{=\frac{1}{c} \\ \text{(since } n!=c\text{)}}} \underbrace{m(m-1)\cdots(m-n+1)}_{=\prod_{i=0}^{n-1}(m-i)}$$
$$= \frac{1}{c}\prod_{i=0}^{n-1}(m-i). \tag{674}$$

Applying this to $m = sa$, we obtain

$$\dbinom{sa}{n} = \frac{1}{c}\prod_{i=0}^{n-1}(sa-i) = \frac{\prod_{i=0}^{n-1}(sa-i)}{c}.$$

Thus,

$$\frac{\prod_{i=0}^{n-1}(sa-i)}{c} = \dbinom{sa}{n} \in \mathbb{Z}$$

(by (238) (applied to $m = sa$)). In other words, $c \mid \prod_{i=0}^{n-1}(sa-i)$. In other words,

$$\prod_{i=0}^{n-1}(sa-i) \equiv 0 \mod c. \tag{675}$$

Now, (673) becomes

$$b^{c-1}\prod_{i=0}^{n-1}(a-bi) \equiv b^{c-1}\underbrace{\prod_{i=0}^{n-1}(b(sa-i))}_{=b^n\prod_{i=0}^{n-1}(sa-i)} = b^{c-1}b^n \underbrace{\prod_{i=0}^{n-1}(sa-i)}_{\substack{\equiv 0 \mod c \\ \text{(by (675))}}} \equiv 0 \mod c.$$

In other words,

$$c \mid b^{c-1}\prod_{i=0}^{n-1}(a-bi). \tag{676}$$

But (674) (applied to $m = a/b$) yields

$$\dbinom{a/b}{n} = \frac{1}{c}\prod_{i=0}^{n-1}\underbrace{(a/b-i)}_{=\frac{1}{b}(a-bi)} = \frac{1}{c}\underbrace{\prod_{i=0}^{n-1}\left(\frac{1}{b}(a-bi)\right)}_{=\left(\frac{1}{b}\right)^n\prod_{i=0}^{n-1}(a-bi)} = \frac{1}{c}\underbrace{\left(\frac{1}{b}\right)^n}_{=\frac{1}{b^n}}\prod_{i=0}^{n-1}(a-bi)$$
$$= \frac{1}{c}\cdot\frac{1}{b^n}\prod_{i=0}^{n-1}(a-bi).$$

Multiplying this equality with $b^{n+(c-1)}$, we obtain

$$b^{n+(c-1)}\binom{a/b}{n} = b^{n+(c-1)} \cdot \frac{1}{c} \cdot \frac{1}{b^n} \prod_{i=0}^{n-1} (a-bi) = \frac{1}{c} \cdot \underbrace{b^{n+(c-1)} \cdot \frac{1}{b^n}}_{=b^{c-1}} \prod_{i=0}^{n-1} (a-bi)$$
$$= \frac{1}{c} \cdot b^{c-1} \prod_{i=0}^{n-1} (a-bi) = \frac{b^{c-1} \prod_{i=0}^{n-1} (a-bi)}{c} \in \mathbb{Z}$$

(by (676)). Hence, there exists some $N \in \mathbb{N}$ such that $b^N \binom{a/b}{n} \in \mathbb{Z}$ (namely, $N = n + (c-1)$). This solves Exercise 3.26. $\square$

Remark 7.58. Our above solution of Exercise 3.26 shows that the N in this exercise can be taken to be $n + (n! - 1)$. This is, however, far from being the best possible value of N. A much better value that also works is $\max\{0, 2n-1\}$. Proving this, however, would require a different idea. The second solution below gives a proof of this better value. (Alternatively, this better value can be obtained by studying the exponents of primes appearing in $n!$.)

7.34.2. Second solution

We shall now prepare to give a second solution to Exercise 3.26. Our main goal is to prove the following fact:

Theorem 7.59. Let a and b be two integers such that $b \neq 0$. Let n be a positive integer. Then, $b^{2n-1}\binom{a/b}{n} \in \mathbb{Z}$.

Clearly, Exercise 3.26 immediately follows from Theorem 7.59 in the case when $n \neq 0$. (In the case when $n = 0$, it holds for obvious reasons.)

Before we start proving Theorem 7.59, let us show the following lemma:

Lemma 7.60. Let b and n be positive integers. Assume that every $k \in \{1, 2, \ldots, n-1\}$ and $c \in \mathbb{Z}$ satisfy

$$b^{2k-1}\binom{c/b}{k} \in \mathbb{Z}. \tag{677}$$

Then:

(a) Every $u \in \mathbb{Z}$ and $v \in \mathbb{Z}$ satisfy

$$b^{2n-2}\left(\binom{(u+v)/b}{n} - \binom{u/b}{n} - \binom{v/b}{n}\right) \in \mathbb{Z}.$$

(b) Every $a \in \mathbb{Z}$ and $h \in \mathbb{N}$ satisfy

$$b^{2n-2}\left(\binom{ha/b}{n} - h\binom{a/b}{n}\right) \in \mathbb{Z}.$$

(c) Every $a \in \mathbb{Z}$ satisfies

$$b^{2n-1}\binom{a/b}{n} \in \mathbb{Z}.$$

Proof of Lemma 7.60. **(a)** Let $u \in \mathbb{Z}$ and $v \in \mathbb{Z}$. But Proposition 3.3 **(a)** (applied to u/b instead of m) yields $\binom{u/b}{0} = 1$. Similarly, $\binom{v/b}{0} = 1$.

Theorem 3.29 (applied to u/b and v/b instead of x and y) yields

$$\begin{aligned}\binom{u/b+v/b}{n} &= \sum_{k=0}^{n}\binom{u/b}{k}\binom{v/b}{n-k} \\ &= \underbrace{\binom{u/b}{0}}_{=1}\underbrace{\binom{v/b}{n-0}}_{=\binom{v/b}{n}} + \sum_{k=1}^{n-1}\binom{u/b}{k}\binom{v/b}{n-k} + \binom{u/b}{n}\underbrace{\binom{v/b}{n-n}}_{=\binom{v/b}{0}=1} \\ &\qquad \left(\begin{array}{c}\text{here, we have split off the addends for } k=0 \text{ and} \\ \text{for } k=n \text{ from the sum (since 0 and } n \text{ are two distinct} \\ \text{elements of } \{0,1,\ldots,n\}\text{)}\end{array}\right) \\ &= \binom{v/b}{n} + \sum_{k=1}^{n-1}\binom{u/b}{k}\binom{v/b}{n-k} + \binom{u/b}{n}.\end{aligned}$$

Subtracting $\binom{u/b}{n} + \binom{v/b}{n}$ from this equality, we obtain

$$\binom{u/b+v/b}{n} - \binom{u/b}{n} - \binom{v/b}{n} = \sum_{k=1}^{n-1}\binom{u/b}{k}\binom{v/b}{n-k}. \tag{678}$$

But for every $k \in \{1,2,\ldots,n-1\}$, the number $b^{2n-2}\binom{u/b}{k}\binom{v/b}{n-k}$ is an integer[332]. Hence, $\sum_{k=1}^{n-1} b^{2n-2}\binom{u/b}{k}\binom{v/b}{n-k}$ is a sum of $n-1$ integers, and thus itself is

[332] *Proof:* Let $k \in \{1,2,\ldots,n-1\}$.

From (677) (applied to $c = u$), we obtain $b^{2k-1}\binom{u/b}{k} \in \mathbb{Z}$. In other words, the number $b^{2k-1}\binom{u/b}{k}$ is an integer.

an integer. In other words,

$$\sum_{k=1}^{n-1} b^{2n-2} \binom{u/b}{k} \binom{v/b}{n-k} \in \mathbb{Z}. \tag{679}$$

Now,

$$b^{2n-2} \left(\underbrace{\binom{(u+v)/b}{n}}_{\substack{=\binom{u/b+v/b}{n} \\ \text{(since } (u+v)/b = u/b+v/b)}} - \binom{u/b}{n} - \binom{v/b}{n} \right)$$
$$= b^{2n-2} \underbrace{\left(\binom{u/b+v/b}{n} - \binom{u/b}{n} - \binom{v/b}{n} \right)}_{\substack{= \sum_{k=1}^{n-1} \binom{u/b}{k}\binom{v/b}{n-k} \\ \text{(by (678))}}}$$
$$= b^{2n-2} \sum_{k=1}^{n-1} \binom{u/b}{k} \binom{v/b}{n-k} = \sum_{k=1}^{n-1} b^{2n-2} \binom{u/b}{k} \binom{v/b}{n-k} \in \mathbb{Z}$$

(by (679)). This proves Lemma 7.60 **(a)**.

(b) Let $a \in \mathbb{Z}$ and $h \in \mathbb{N}$. For every $k \in \{0, 1, \ldots, h-1\}$, the number

---

But we also have $n-k \in \{1, 2, \ldots, n-1\}$ (since $k \in \{1, 2, \ldots, n-1\}$). Thus, from (677) (applied to $n-k$ and v instead of k and c), we obtain $b^{2(n-k)-1} \binom{v/b}{n-k} \in \mathbb{Z}$. In other words, the number $b^{2(n-k)-1} \binom{v/b}{n-k}$ is an integer.

Now, the numbers $b^{2k-1} \binom{u/b}{k}$ and $b^{2(n-k)-1} \binom{v/b}{n-k}$ are integers. Hence, their product is an integer as well. In other words, the number $b^{2k-1} \binom{u/b}{k} \cdot b^{2(n-k)-1} \binom{v/b}{n-k}$ is an integer. Since

$$b^{2k-1} \binom{u/b}{k} \cdot b^{2(n-k)-1} \binom{v/b}{n-k}$$
$$= \underbrace{b^{2k-1} b^{2(n-k)-1}}_{\substack{= b^{(2k-1)+(2(n-k)-1)} = b^{2n-2} \\ \text{(since } (2k-1)+(2(n-k)-1) = 2n-2)}} \binom{u/b}{k} \binom{v/b}{n-k} = b^{2n-2} \binom{u/b}{k} \binom{v/b}{n-k},$$

this rewrites as follows: The number $b^{2n-2} \binom{u/b}{k} \binom{v/b}{n-k}$ is an integer. Qed.

$b^{2n-2}\left(\dbinom{(k+1)\,a/b}{n}-\dbinom{ka/b}{n}-\dbinom{a/b}{n}\right)$ is an integer[333]. Hence,

$$\sum_{k=0}^{h-1} b^{2n-2}\left(\dbinom{(k+1)\,a/b}{n}-\dbinom{ka/b}{n}-\dbinom{a/b}{n}\right)$$

is a sum of h integers, and thus itself is an integer. In other words,

$$\sum_{k=0}^{h-1} b^{2n-2}\left(\dbinom{(k+1)\,a/b}{n}-\dbinom{ka/b}{n}-\dbinom{a/b}{n}\right)\in\mathbb{Z}. \tag{680}$$

But $1-1=0\leq h$. Hence, (16) (applied to $\mathbb{A}=\mathbb{Q}$, $u=1$, $v=h$ and $a_s=\dbinom{sa/b}{n}$) yields

$$\sum_{s=1}^{h}\left(\dbinom{sa/b}{n}-\dbinom{(s-1)\,a/b}{n}\right)=\dbinom{ha/b}{n}-\underbrace{\dbinom{(1-1)\,a/b}{n}}_{\substack{=\dbinom{0}{n}\\ (\text{since } (1-1)a/b=0)}}$$
$$=\dbinom{ha/b}{n}-\underbrace{\dbinom{0}{n}}_{\substack{=0\\ (\text{by Proposition 3.6,}\\ \text{applied to } m=0)}}=\dbinom{ha/b}{n}.$$

Hence,

$$\dbinom{ha/b}{n}=\sum_{s=1}^{h}\left(\dbinom{sa/b}{n}-\dbinom{(s-1)\,a/b}{n}\right)$$
$$=\sum_{k=0}^{h-1}\left(\dbinom{(k+1)\,a/b}{n}-\dbinom{ka/b}{n}\right) \tag{681}$$

[333] *Proof:* Let $k\in\{0,1,\ldots,h-1\}$. Then, Lemma 7.60 **(a)** (applied to $u=ka$ and $v=a$) yields

$$b^{2n-2}\left(\dbinom{(ka+a)\,/b}{n}-\dbinom{ka/b}{n}-\dbinom{a/b}{n}\right)\in\mathbb{Z}.$$

Since $ka+a=(k+1)\,a$, this rewrites as follows:

$$b^{2n-2}\left(\dbinom{(k+1)\,a/b}{n}-\dbinom{ka/b}{n}-\dbinom{a/b}{n}\right)\in\mathbb{Z}.$$

Qed.

(here, we have substituted $k+1$ for s in the sum). Now,

$$\sum_{k=0}^{h-1}\left(\binom{(k+1)a/b}{n}-\binom{ka/b}{n}-\binom{a/b}{n}\right)$$
$$=\underbrace{\sum_{k=0}^{h-1}\left(\binom{(k+1)a/b}{n}-\binom{ka/b}{n}\right)}_{\substack{=\binom{ha/b}{n}\\ \text{(by (681))}}}-\underbrace{\sum_{k=0}^{h-1}\binom{a/b}{n}}_{=h\binom{a/b}{n}}$$
$$=\binom{ha/b}{n}-h\binom{a/b}{n}. \qquad (682)$$

Now,

$$\sum_{k=0}^{h-1}b^{2n-2}\left(\binom{(k+1)a/b}{n}-\binom{ka/b}{n}-\binom{a/b}{n}\right)$$
$$=b^{2n-2}\underbrace{\sum_{k=0}^{h-1}\left(\binom{(k+1)a/b}{n}-\binom{ka/b}{n}-\binom{a/b}{n}\right)}_{\substack{=\binom{ha/b}{n}-h\binom{a/b}{n}\\ \text{(by (682))}}}$$
$$=b^{2n-2}\left(\binom{ha/b}{n}-h\binom{a/b}{n}\right).$$

Thus,

$$b^{2n-2}\left(\binom{ha/b}{n}-h\binom{a/b}{n}\right)=\sum_{k=0}^{h-1}b^{2n-2}\left(\binom{(k+1)a/b}{n}-\binom{ka/b}{n}-\binom{a/b}{n}\right)$$
$$\in\mathbb{Z} \qquad \text{(by (680))}.$$

This proves Lemma 7.60 **(b)**.

(c) Let $a\in\mathbb{Z}$. Proposition 3.20 (applied to $m=a$) yields $\binom{a}{n}\in\mathbb{Z}$. In other words, $\binom{a}{n}$ is an integer.

Also, $2n-2\in\mathbb{N}$ (since n is a positive integer). Thus, b^{2n-2} is an integer (since b is an integer). Now, the numbers b^{2n-2} and $\binom{a}{n}$ are both integers. Hence, their product must also be an integer. In other words, $b^{2n-2}\binom{a}{n}$ is an integer.

But Lemma 7.60 **(b)** (applied to $h=b$) yields

$$b^{2n-2}\left(\binom{ba/b}{n}-b\binom{a/b}{n}\right)\in\mathbb{Z}.$$

In other words, $b^{2n-2}\left(\dbinom{ba/b}{n}-b\dbinom{a/b}{n}\right)$ is an integer. Denote this integer by z. Thus,

$$z=b^{2n-2}\left(\dbinom{ba/b}{n}-b\dbinom{a/b}{n}\right)=b^{2n-2}\underbrace{\dbinom{ba/b}{n}}_{=\dbinom{a}{n}}-\underbrace{b^{2n-2}b}_{=b^{(2n-2)+1}=b^{2n-1}}\dbinom{a/b}{n}$$
$$=b^{2n-2}\dbinom{a}{n}-b^{2n-1}\dbinom{a/b}{n}.$$

Solving this equality for $b^{2n-1}\dbinom{a/b}{n}$, we obtain

$$b^{2n-1}\dbinom{a/b}{n}=b^{2n-2}\dbinom{a}{n}-z. \tag{683}$$

But the numbers $b^{2n-2}\dbinom{a}{n}$ and z are integers. Hence, their difference is also an integer. In other words, $b^{2n-2}\dbinom{a}{n}-z$ is an integer. In other words, $b^{2n-2}\dbinom{a}{n}-z\in\mathbb{Z}$. Hence, (683) becomes $b^{2n-1}\dbinom{a/b}{n}=b^{2n-2}\dbinom{a}{n}-z\in\mathbb{Z}$. This proves Lemma 7.60 **(c)**. □

Our next lemma is essentially Theorem 7.59, restricted to the case when b is positive:

Lemma 7.61. Let a and b be two integers such that $b>0$. Let n be a positive integer. Then, $b^{2n-1}\dbinom{a/b}{n}\in\mathbb{Z}$.

Proof of Lemma 7.61. We shall prove Lemma 7.61 by strong induction on n:
Induction step: Let N be a positive integer. Assume that Lemma 7.61 holds in the case when $n<N$. We must show that Lemma 7.61 holds in the case when $n=N$.
We have assumed that Lemma 7.61 holds in the case when $n<N$. In other words, the following statement holds:

Statement 1: Let a and b be two integers such that $b>0$. Let n be a positive integer such that $n<N$. Then, $b^{2n-1}\dbinom{a/b}{n}\in\mathbb{Z}$.

Now, let us prove the following statement:

Statement 2: Let a and b be two integers such that $b>0$. Then, $b^{2N-1}\dbinom{a/b}{N}\in\mathbb{Z}$.

[*Proof of Statement 2:* Every $k \in \{1, 2, \ldots, N-1\}$ and $c \in \mathbb{Z}$ satisfy

$$b^{2k-1} \binom{c/b}{k} \in \mathbb{Z}$$

[334]. Hence, Lemma 7.60 **(c)** (applied to $n = N$) yields $b^{2N-1} \binom{a/b}{N} \in \mathbb{Z}$. Thus, Statement 2 is proven.]

So we have proven Statement 2. In other words, we have proven that Lemma 7.61 holds in the case when $n = N$. This completes the induction step. Thus, the inductive proof of Lemma 7.61 is complete. □

Proof of Theorem 7.59. We must prove that $b^{2n-1} \binom{a/b}{n} \in \mathbb{Z}$. If $b > 0$, then this follows immediately from Lemma 7.61. Hence, for the rest of this proof, we WLOG assume that we don't have $b > 0$. Hence, $b \leq 0$, so that $b < 0$ (since $b \neq 0$). Therefore, $-b > 0$. Thus, Lemma 7.60 (applied to $-a$ and $-b$ instead of a and b) yields $(-b)^{2n-1} \binom{(-a)/(-b)}{n} \in \mathbb{Z}$. Since $(-a)/(-b) = a/b$, this rewrites as $(-b)^{2n-1} \binom{a/b}{n} \in \mathbb{Z}$. In other words, $(-b)^{2n-1} \binom{a/b}{n}$ is an integer.

But

$$\underbrace{(-b)^{2n-1}}_{=(-1)^{2n-1} b^{2n-1}} \binom{a/b}{n} = \underbrace{(-1)^{2n-1}}_{\substack{=-1 \\ \text{(since } 2n-1 \text{ is odd)}}} b^{2n-1} \binom{a/b}{n} = -b^{2n-1} \binom{a/b}{n}.$$

In other words, the two numbers $(-b)^{2n-1} \binom{a/b}{n}$ and $b^{2n-1} \binom{a/b}{n}$ differ only in sign. Since the first of them is an integer, we thus conclude that so is the second. In other words, $b^{2n-1} \binom{a/b}{n} \in \mathbb{Z}$. This proves Theorem 7.59. □

Now, only some trivial bookkeeping remains to be done in order to solve Exercise 3.26. To simplify it, we state the following corollary from Theorem 7.59:

Corollary 7.62. Let a and b be two integers such that $b \neq 0$. Let $n \in \mathbb{N}$. Let $m = \max\{0, 2n-1\}$. Then, $b^m \binom{a/b}{n} \in \mathbb{Z}$.

Proof of Corollary 7.62. If $n = 0$, then Corollary 7.62 holds[335]. Hence, for the rest of this proof, we WLOG assume that $n \neq 0$. Therefore, n is a positive integer

[334] *Proof.* Let $k \in \{1, 2, \ldots, N-1\}$ and $c \in \mathbb{Z}$. From $k \in \{1, 2, \ldots, N-1\}$, we obtain $1 \leq k \leq N-1$. Now, k is a positive integer (since $1 \leq k$) and satisfies $k < N$ (since $k \leq N-1 < N$). Hence, Statement 1 (applied to k and c instead of n and a) yields $b^{2k-1} \binom{c/b}{k} \in \mathbb{Z}$. Qed.

[335] *Proof.* Assume that $n = 0$. We must show that Corollary 7.62 holds.

(since $n \in \mathbb{N}$). Therefore, $2n-1 > 0$, so that $\max\{0, 2n-1\} = 2n-1$. Now, $m = \max\{0, 2n-1\} = 2n-1$, so that $b^m = b^{2n-1}$. Multiplying this equality by $\dbinom{a/b}{n}$, we find

$$b^m \dbinom{a/b}{n} = b^{2n-1} \dbinom{a/b}{n} \in \mathbb{Z}$$

(by Theorem 7.59). Corollary 7.62 is thus proven. □

Second solution to Exercise 3.26. Let $m = \max\{0, 2n-1\}$. Then, $m = \max\{0, 2n-1\} \geq 0$, so that $m \in \mathbb{N}$. Also, Corollary 7.62 shows that $b^m \dbinom{a/b}{n} \in \mathbb{Z}$. Hence, there exists some $N \in \mathbb{N}$ such that $b^N \dbinom{a/b}{n} \in \mathbb{Z}$ (namely, $N = m$). This solves Exercise 3.26. □

We remark that Theorem 7.59 is equivalent to [AndDos12, §3.3, problem 4] (which is a problem from the IMO Shortlist 1985). Indeed, the latter problem claims the following:

Theorem 7.63. Let a and b be integers. Let n be a positive integer. Then,

$$\frac{1}{n!} \cdot a(a+b)(a+2b)\cdots(a+(n-1)b) \cdot b^{n-1} \in \mathbb{Z}.$$

Let us derive Theorem 7.63 from Theorem 7.59:

Proof of Theorem 7.63. We are in one of the following three cases:

Case 1: We have $b \neq 0$.

Case 2: We have $n = 1$.

Case 3: We have neither $b \neq 0$ nor $n = 1$.

Let us first consider Case 1. In this case, we have $b \neq 0$. Thus, $-b \neq 0$ as well. Hence, Theorem 7.59 (applied to $-b$ instead of b) yields $(-b)^{2n-1} \dbinom{a/(-b)}{n} \in \mathbb{Z}$.

We have $m = \max\{0, 2n-1\} = 0$ (since $2\underbrace{n}_{=0} - 1 = 0 - 1 < 0$). Hence, $b^m = b^0 = 1$ and thus

$$\underbrace{b^m}_{=1} \dbinom{a/b}{n} = \dbinom{a/b}{n} = \dbinom{a/b}{0} \qquad (\text{since } n = 0).$$

But Proposition 3.3 **(a)** (applied to a/b instead of m) yields $\dbinom{a/b}{0} = 1$. Thus, $b^m \dbinom{a/b}{n} = \dbinom{a/b}{0} = 1 \in \mathbb{Z}$. Thus, Corollary 7.62 holds.

But (226) (applied to $m = a/(-b)$) yields

$$\begin{aligned}
\binom{a/(-b)}{n} &= \frac{(a/(-b))(a/(-b)-1)\cdots(a/(-b)-n+1)}{n!} \\
&= \frac{1}{n!}\cdot \underbrace{(a/(-b))(a/(-b)-1)\cdots(a/(-b)-n+1)}_{=\prod_{i=0}^{n-1}(a/(-b)-i)} \\
&= \frac{1}{n!}\cdot \prod_{i=0}^{n-1} \underbrace{(a/(-b)-i)}_{=\frac{a+ib}{-b}} = \frac{1}{n!}\cdot \underbrace{\prod_{i=0}^{n-1}\frac{a+ib}{-b}}_{=\frac{\prod_{i=0}^{n-1}(a+ib)}{(-b)^n}} \\
&= \frac{1}{n!}\cdot \frac{\prod_{i=0}^{n-1}(a+ib)}{(-b)^n}.
\end{aligned}$$

Multiplying both sides of this equality by $(-b)^{2n-1}$, we obtain

$$\begin{aligned}
(-b)^{2n-1}\binom{a/(-b)}{n} &= (-b)^{2n-1}\frac{1}{n!}\cdot\frac{\prod_{i=0}^{n-1}(a+ib)}{(-b)^n} = \frac{1}{n!}\cdot\underbrace{\frac{(-b)^{2n-1}}{(-b)^n}}_{=(-b)^{n-1}}\cdot\prod_{i=0}^{n-1}(a+ib) \\
&= \frac{1}{n!}\cdot(-b)^{n-1}\cdot\prod_{i=0}^{n-1}(a+ib).
\end{aligned}$$

Multiplying both sides of this equality by $(-1)^{n-1}$, we obtain

$$\begin{aligned}
&(-1)^{n-1}(-b)^{2n-1}\binom{a/(-b)}{n} \\
&= (-1)^{n-1}\frac{1}{n!}\cdot(-b)^{n-1}\cdot\prod_{i=0}^{n-1}(a+ib) = \frac{1}{n!}\cdot\underbrace{(-1)^{n-1}(-b)^{n-1}}_{=((-1)(-b))^{n-1}}\cdot\prod_{i=0}^{n-1}(a+ib) \\
&= \frac{1}{n!}\cdot\Bigg(\underbrace{(-1)(-b)}_{=b}\Bigg)^{n-1}\cdot\prod_{i=0}^{n-1}(a+ib) = \frac{1}{n!}\cdot b^{n-1}\cdot\prod_{i=0}^{n-1}(a+ib) \\
&= \frac{1}{n!}\cdot\underbrace{\prod_{i=0}^{n-1}(a+ib)}_{=a(a+b)(a+2b)\cdots(a+(n-1)b)}\cdot b^{n-1} \\
&= \frac{1}{n!}\cdot a(a+b)(a+2b)\cdots(a+(n-1)b)\cdot b^{n-1}.
\end{aligned}$$

Hence,

$$\begin{aligned}&\frac{1}{n!}\cdot a\left(a+b\right)\left(a+2b\right)\cdots\left(a+\left(n-1\right)b\right)\cdot b^{n-1}\\&=\underbrace{\left(-1\right)^{n-1}}_{\in\mathbb{Z}}\underbrace{\left(-b\right)^{2n-1}\binom{a/\left(-b\right)}{n}}_{\in\mathbb{Z}}\in\mathbb{Z}.\end{aligned}$$

Hence, Theorem 7.63 is proven in Case 1.

Let us now consider Case 2. In this case, we have $n=1$. Thus, $\frac{1}{n!}=\frac{1}{1!}=\frac{1}{1}=1\in\mathbb{Z}$ and $b^{n-1}=b^{1-1}=b^0=1\in\mathbb{Z}$, and therefore

$$\underbrace{\frac{1}{n!}}_{\in\mathbb{Z}}\cdot\underbrace{a\left(a+b\right)\left(a+2b\right)\cdots\left(a+\left(n-1\right)b\right)}_{\substack{\in\mathbb{Z}\\ \text{(since } a \text{ and } b \text{ are integers)}}}\cdot\underbrace{b^{n-1}}_{\in\mathbb{Z}}\in\mathbb{Z}.$$

Hence, Theorem 7.63 is proven in Case 2.

Let us finally consider Case 3. In this case, we have neither $b\neq 0$ nor $n=1$. Thus, $b=0$ (since we don't have $b\neq 0$) and $n\neq 1$ (since we don't have $n=1$). From $n\neq 1$, we obtain $n\geq 2$ (since n is a positive integer), and thus $n-1\geq 1$. Hence, $0^{n-1}=0$. But from $b=0$, we obtain $b^{n-1}=0^{n-1}=0$. Thus,

$$\frac{1}{n!}\cdot a\left(a+b\right)\left(a+2b\right)\cdots\left(a+\left(n-1\right)b\right)\cdot\underbrace{b^{n-1}}_{=0}=0\in\mathbb{Z}.$$

Hence, Theorem 7.63 is proven in Case 3.

We have now proven Theorem 7.63 in each of the three Cases 1, 2 and 3. Since these three Cases cover all possibilities, we thus conclude that Theorem 7.63 always holds. □

Similarly, we can derive Theorem 7.59 back from Theorem 7.63. (That derivation is actually easier, since we don't need to distinguish between any cases.)

7.35. Solution to Exercise 3.27

We shall derive Exercise 3.27 from a sequence of lemmas. The first one will be an easy consequence from the binomial identity:

Lemma 7.64. Let $g\in\mathbb{N}$. Let x and y be two real numbers. Let $z=x+y$.

(a) Then,

$$z^g=\sum_{i=0}^{g}\binom{g}{i}x^iy^{g-i}.$$

(b) Let $n\in\mathbb{N}$ be such that $n\geq g$. Then,

$$y^{n-g}z^g=\sum_{i=0}^{n}\binom{g}{i}x^iy^{n-i}.$$

Proof of Lemma 7.64. Proposition 3.21 (applied to g instead of n) yields

$$(x+y)^g = \sum_{k=0}^{g} \binom{g}{k} x^k y^{g-k} = \sum_{i=0}^{g} \binom{g}{i} x^i y^{g-i}$$

(here, we have renamed the summation index k as i in the sum). Now, from $z = x+y$, we obtain

$$z^g = (x+y)^g = \sum_{i=0}^{g} \binom{g}{i} x^i y^{g-i}. \qquad (684)$$

This proves Lemma 7.64 **(a)**.

(b) We have $n-g \geq 0$ (since $n \geq g$), and thus $n-g \in \mathbb{N}$. Hence, y^{n-g} is well-defined.

Multiplying both sides of the equality (684) by y^{n-g}, we obtain

$$y^{n-g}z^g = y^{n-g} \sum_{i=0}^{g} \binom{g}{i} x^i y^{g-i} = \sum_{i=0}^{g} \binom{g}{i} x^i \underbrace{y^{n-g}y^{g-i}}_{\substack{=y^{(n-g)+(g-i)}=y^{n-i} \\ (\text{since } (n-g)+(g-i)=n-i)}}$$
$$= \sum_{i=0}^{g} \binom{g}{i} x^i y^{n-i}. \qquad (685)$$

On the other hand, for each $i \in \{g+1, g+2, \ldots, n\}$, we have

$$\binom{g}{i} = 0 \qquad (686)$$

[336].

But $g \geq 0$ (since $g \in \mathbb{N}$), so that $0 \leq g \leq n$ (since $n \geq g$). Hence, we can split the sum $\sum_{i=0}^{n} \binom{g}{i} x^i y^{n-i}$ as follows:

$$\sum_{i=0}^{n} \binom{g}{i} x^i y^{n-i} = \sum_{i=0}^{g} \binom{g}{i} x^i y^{n-i} + \sum_{i=g+1}^{n} \underbrace{\binom{g}{i}}_{\substack{=0 \\ (\text{by } (686))}} x^i y^{n-i}$$
$$= \sum_{i=0}^{g} \binom{g}{i} x^i y^{n-i} + \underbrace{\sum_{i=g+1}^{n} 0 x^i y^{n-i}}_{=0} = \sum_{i=0}^{g} \binom{g}{i} x^i y^{n-i}.$$

Comparing this with (685), we obtain $y^{n-g}z^g = \sum_{i=0}^{n} \binom{g}{i} x^i y^{n-i}$. This proves Lemma 7.64 **(b)**. $\square$

[336] *Proof of (686):* Let $i \in \{g+1, g+2, \ldots, n\}$. Thus, $i \geq g+1 > g$, so that $g < i$. Also, $i > g \geq 0$ (since $g \in \mathbb{N}$), so that $i \in \mathbb{N}$. Hence, Proposition 3.6 (applied to g and i instead of m and n) yields $\binom{g}{i} = 0$. This proves (686).

Our next lemma is a combinatorial identity that follows easily from Proposition 3.32 **(f)**:

Lemma 7.65. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $i \in \{0, 1, \ldots, n\}$. Then,

$$\sum_{k=0}^{n} \binom{m+k}{k} \binom{n-k}{i} = \binom{n+m+1}{m+i+1}.$$

Proof of Lemma 7.65. We have $i \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$. Also, from $i \in \{0, 1, \ldots, n\}$, we obtain $i \leq n$. Hence, $m + \underbrace{i}_{\leq n} + 1 \leq m + n + 1 = n + m + 1$. In other words, $n + m + 1 \geq m + i + 1$. Also, $m + i + 1 \in \mathbb{N}$ (since $m \in \mathbb{N}$ and $i \in \mathbb{N}$) and $n + m + 1 \in \mathbb{N}$ (since $n \in \mathbb{N}$ and $m \in \mathbb{N}$). Thus, Proposition 3.8 (applied to $n + m + 1$ and $m + i + 1$ instead of m and n) yields

$$\binom{n+m+1}{m+i+1} = \binom{n+m+1}{(n+m+1)-(m+i+1)} = \binom{n+m+1}{n-i} \tag{687}$$

(since $(n + m + 1) - (m + i + 1) = n - i$).

For each $k \in \mathbb{N}$, we have

$$\binom{m+k}{m} = \binom{m+k}{k} \tag{688}$$

(by Lemma 7.54 (applied to k instead of n)).

For each $k \in \{0, 1, \ldots, m-1\}$, we have

$$\binom{k}{m} = 0 \tag{689}$$

[337].

We have $m \geq 0$ (since $m \in \mathbb{N}$) and $n \geq 0$ (since $n \in \mathbb{N}$), so that $\underbrace{n}_{\geq 0} + m \geq m$ and thus $m \leq n + m$. From $m \geq 0$, we obtain $0 \leq m \leq n + m$. Hence, we can split the

[337] *Proof of (689):* Let $k \in \{0, 1, \ldots, m-1\}$. Thus, $k \geq 0$ and $k \leq m - 1 < m$. From $k \geq 0$, we obtain $k \in \mathbb{N}$. Hence, Proposition 3.6 (applied to k and m instead of m and n) yields $\binom{k}{m} = 0$. This proves (689).

sum $\sum\limits_{k=0}^{n+m} \dbinom{k}{m} \dbinom{n+m-k}{i}$ as follows:

$$\begin{aligned}
&\sum_{k=0}^{n+m} \binom{k}{m} \binom{n+m-k}{i} \\
&= \sum_{k=0}^{m-1} \underbrace{\binom{k}{m}}_{\substack{=0 \\ \text{(by (689))}}} \binom{n+m-k}{i} + \sum_{k=m}^{n+m} \binom{k}{m} \binom{n+m-k}{i} \\
&= \underbrace{\sum_{k=0}^{m-1} 0 \binom{n+m-k}{i}}_{=0} + \sum_{k=m}^{n+m} \binom{k}{m} \binom{n+m-k}{i} \\
&= \sum_{k=m}^{n+m} \binom{k}{m} \binom{n+m-k}{i} = \sum_{k=0}^{n} \underbrace{\binom{k+m}{m}}_{\substack{=\binom{m+k}{m} \\ \text{(since } k+m=m+k)}} \underbrace{\binom{n+m-(k+m)}{i}}_{\substack{=\binom{n-k}{i} \\ \text{(since } n+m-(k+m)=n-k)}} \\
&\qquad \text{(here, we have substituted } k+m \text{ for } k \text{ in the sum)} \\
&= \sum_{k=0}^{n} \underbrace{\binom{m+k}{m}}_{\substack{=\binom{m+k}{k} \\ \text{(by (688))}}} \binom{n-k}{i} = \sum_{k=0}^{n} \binom{m+k}{k} \binom{n-k}{i}. \qquad (690)
\end{aligned}$$

But Proposition 3.32 **(f)** (applied to $n+m$, m and i instead of n, x and y) yields

$$\binom{n+m+1}{m+i+1} = \sum_{k=0}^{n+m} \binom{k}{m} \binom{n+m-k}{i} = \sum_{k=0}^{n} \binom{m+k}{k} \binom{n-k}{i}$$

(by (690)). This proves Lemma 7.65. □

Our next step brings us much closer to Exercise 3.27:

Proposition 7.66. Let x and y be two real numbers. Let $z = x + y$. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then:

(a) We have

$$x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k z^{n-k} = \sum_{i=m+1}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i}.$$

(b) We have

$$y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k z^{m-k} = \sum_{i=0}^{m} \binom{n+m+1}{i} x^i y^{(n+m+1)-i}.$$

(c) We have

$$x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k z^{n-k} + y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k z^{m-k} = z^{n+m+1}.$$

Proof of Proposition 7.66. **(a)** Let $k \in \{0, 1, \ldots, n\}$. Thus, $k \leq n$, so that $n - k \geq 0$ and thus $n - k \in \mathbb{N}$. Also, $k \geq 0$ (since $k \in \{0, 1, \ldots, n\}$), so that $n - \underbrace{k}_{\geq 0} \leq n$ and thus $n \geq n - k$. Hence, Lemma 7.64 **(a)** (applied to $g = n - k$) yields

$$y^{n-(n-k)} z^{n-k} = \sum_{i=0}^{n} \binom{n-k}{i} x^i y^{n-i}.$$

In view of $n - (n - k) = k$, this rewrites as follows:

$$y^k z^{n-k} = \sum_{i=0}^{n} \binom{n-k}{i} x^i y^{n-i}. \qquad (691)$$

Now, forget that we fixed k. We thus have proven the equality (691) for each

$k \in \{0, 1, \ldots, n\}$. Now,

$$x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} \underbrace{y^k z^{n-k}}_{\substack{= \sum_{i=0}^{n} \binom{n-k}{i} x^i y^{n-i} \\ \text{(by (691))}}}$$

$$= x^{m+1} \underbrace{\sum_{k=0}^{n} \binom{m+k}{k} \sum_{i=0}^{n} \binom{n-k}{i} x^i y^{n-i}}_{= \sum_{i=0}^{n} \binom{m+k}{k} \binom{n-k}{i} x^i y^{n-i}} = x^{m+1} \underbrace{\sum_{k=0}^{n} \sum_{i=0}^{n}}_{= \sum_{i=0}^{n} \sum_{k=0}^{n}} \binom{m+k}{k} \binom{n-k}{i} x^i y^{n-i}$$

$$= x^{m+1} \sum_{i=0}^{n} \underbrace{\sum_{k=0}^{n} \binom{m+k}{k} \binom{n-k}{i} x^i y^{n-i}}_{= \left(\sum_{k=0}^{n} \binom{m+k}{k} \binom{n-k}{i} \right) x^i y^{n-i}} = x^{m+1} \sum_{i=0}^{n} \underbrace{\left(\sum_{k=0}^{n} \binom{m+k}{k} \binom{n-k}{i} \right)}_{\substack{= \binom{n+m+1}{m+i+1} \\ \text{(by Lemma 7.65)}}} x^i y^{n-i}$$

$$= x^{m+1} \sum_{i=0}^{n} \underbrace{\binom{n+m+1}{m+i+1}}_{\substack{= \binom{n+m+1}{(m+1)+i} \\ \text{(since } m+i+1=(m+1)+i)}} x^i \underbrace{y^{n-i}}_{\substack{= y^{(n+m+1)-((m+1)+i)} \\ \text{(since } n-i=(n+m+1)-((m+1)+i))}}$$

$$= x^{m+1} \sum_{i=0}^{n} \binom{n+m+1}{(m+1)+i} x^i y^{(n+m+1)-((m+1)+i)}$$

$$= \sum_{i=0}^{n} \binom{n+m+1}{(m+1)+i} \underbrace{x^{m+1} x^i}_{= x^{(m+1)+i}} y^{(n+m+1)-((m+1)+i)}$$

$$= \sum_{i=0}^{n} \binom{n+m+1}{(m+1)+i} x^{(m+1)+i} y^{(n+m+1)-((m+1)+i)}$$

$$= \sum_{i=m+1}^{(m+1)+n} \binom{n+m+1}{i} x^i y^{(n+m+1)-i}$$

(here, we have substituted i for $(m+1)+i$ in the sum)

$$= \sum_{i=m+1}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} \qquad (\text{since } (m+1)+n = n+m+1).$$

This proves Proposition 7.66 **(a)**.

(b) Each $i \in \{n+1, n+2, \ldots, n+m+1\}$ satisfies

$$\binom{n+m+1}{i} = \binom{n+m+1}{(n+m+1)-i} \qquad (692)$$

[338].

We have $z = x + y = y + x$. Hence, Proposition 7.66 **(a)** (applied to y, x, m and n instead of x, y, n and m) yields

$$\begin{aligned}
& y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k z^{m-k} \\
&= \sum_{i=n+1}^{m+n+1} \binom{m+n+1}{i} y^i x^{(m+n+1)-i} \\
&= \sum_{i=n+1}^{n+m+1} \underbrace{\binom{n+m+1}{i}}_{\substack{=\binom{n+m+1}{(n+m+1)-i} \\ \text{(by (692))}}} \underbrace{y^i}_{\substack{=y^{(n+m+1)-((n+m+1)-i)} \\ \text{(since } i=(n+m+1)-((n+m+1)-i))}} x^{(n+m+1)-i} \\
&\qquad \left(\text{since } m+n = n+m\right) \\
&= \sum_{i=n+1}^{n+m+1} \binom{n+m+1}{(n+m+1)-i} y^{(n+m+1)-((n+m+1)-i)} x^{(n+m+1)-i} \\
&= \underbrace{\sum_{i=0}^{(n+m+1)-(n+1)}}_{\substack{=\sum_{i=0}^{m} \\ \text{(since } (n+m+1)-(n+1)=m)}} \binom{n+m+1}{i} \underbrace{y^{(n+m+1)-i} x^i}_{=x^i y^{(n+m+1)-i}} \\
&\qquad \left(\text{here, we have substituted } i \text{ for } (n+m+1)-i \text{ in the sum}\right) \\
&= \sum_{i=0}^{m} \binom{n+m+1}{i} x^i y^{(n+m+1)-i}.
\end{aligned}$$

This proves Proposition 7.66 **(b)**.

(c) We have $\underbrace{n}_{\substack{\geq 0 \\ \text{(since } n\in\mathbb{N})}} + m + 1 \geq m + 1 \geq m$ and thus $m \leq n + m + 1$. Also, from $m \in \mathbb{N}$, we obtain $m \geq 0$, thus $0 \leq m \leq n + m + 1$.

We have $n + m + 1 \in \mathbb{N}$ (since $n \in \mathbb{N}$ and $m \in \mathbb{N}$). Hence, Lemma 7.64 **(a)**

[338] *Proof of (692):* Let $i \in \{n+1, n+2, \ldots, n+m+1\}$. Thus, $i \geq n+1 > n \geq 0$ (since $n \in \mathbb{N}$). Hence, $i \in \mathbb{N}$. Also, $i \in \{n+1, n+2, \ldots, n+m+1\}$ shows that $i \leq n+m+1$, so that $n+m+1 \geq i$. Thus, Proposition 3.8 (applied to $n+m+1$ and i instead of m and n) yields $\binom{n+m+1}{i} = \binom{n+m+1}{(n+m+1)-i}$. This proves (692).

(applied to $n+m+1$ instead of g) yields

$$\begin{aligned}
&z^{n+m+1} \\
&= \sum_{i=0}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} \\
&= \sum_{i=0}^{m} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} + \sum_{i=m+1}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} \qquad (693)
\end{aligned}$$

(here, we have split the sum at $i=m$, since $0 \leq m \leq n+m+1$).

Proposition 7.66 **(a)** yields

$$x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k z^{n-k} = \sum_{i=m+1}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i}. \qquad (694)$$

Proposition 7.66 **(b)** yields

$$y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k z^{m-k} = \sum_{i=0}^{m} \binom{n+m+1}{i} x^i y^{(n+m+1)-i}.$$

Adding this equality to (694), we find

$$\begin{aligned}
&x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k z^{n-k} + y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k z^{m-k} \\
&= \sum_{i=m+1}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} + \sum_{i=0}^{m} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} \\
&= \sum_{i=0}^{m} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} + \sum_{i=m+1}^{n+m+1} \binom{n+m+1}{i} x^i y^{(n+m+1)-i} \\
&= z^{n+m+1} \qquad (\text{by } (693)).
\end{aligned}$$

This proves Proposition 7.66 **(c)**. □

Solution to Exercise 3.27. **(a)** Let x and y be two real numbers such that $x+y=1$. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Thus, $1 = x+y$. Hence, Proposition 7.66 **(c)** (applied to $z=1$) yields

$$x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k 1^{n-k} + y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k 1^{m-k} = 1^{n+m+1} = 1.$$

Comparing this with

$$\begin{aligned}
&x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k \underbrace{1^{n-k}}_{=1} + y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k \underbrace{1^{m-k}}_{=1} \\
&= x^{m+1} \sum_{k=0}^{n} \binom{m+k}{k} y^k + y^{n+1} \sum_{k=0}^{m} \binom{n+k}{k} x^k,
\end{aligned}$$

we obtain

$$x^{m+1}\sum_{k=0}^{n}\binom{m+k}{k}y^k+y^{n+1}\sum_{k=0}^{m}\binom{n+k}{k}x^k=1.$$

This solves Exercise 3.27 **(a)**.

(b) The real numbers $\frac{1}{2}$ and $\frac{1}{2}$ satisfy $\frac{1}{2}+\frac{1}{2}=1$. Hence, Exercise 3.27 **(a)** (applied to $m=n$, $x=\frac{1}{2}$ and $y=\frac{1}{2}$) yields

$$\left(\frac{1}{2}\right)^{n+1}\sum_{k=0}^{n}\binom{n+k}{k}\left(\frac{1}{2}\right)^k+\left(\frac{1}{2}\right)^{n+1}\sum_{k=0}^{n}\binom{n+k}{k}\left(\frac{1}{2}\right)^k=1.$$

Hence,

$$\begin{aligned}1&=\left(\frac{1}{2}\right)^{n+1}\sum_{k=0}^{n}\binom{n+k}{k}\left(\frac{1}{2}\right)^k+\left(\frac{1}{2}\right)^{n+1}\sum_{k=0}^{n}\binom{n+k}{k}\left(\frac{1}{2}\right)^k\\&=2\cdot\underbrace{\left(\frac{1}{2}\right)^{n+1}}_{=\frac{1}{2^{n+1}}}\sum_{k=0}^{n}\binom{n+k}{k}\underbrace{\left(\frac{1}{2}\right)^k}_{=\frac{1}{2^k}}=\underbrace{2\cdot\frac{1}{2^{n+1}}}_{=\frac{1}{2^n}}\sum_{k=0}^{n}\binom{n+k}{k}\frac{1}{2^k}=\frac{1}{2^n}\sum_{k=0}^{n}\binom{n+k}{k}\frac{1}{2^k}.\end{aligned}$$

Multiplying both sides of this equality by 2^n, we obtain $2^n=\sum_{k=0}^{n}\binom{n+k}{k}\frac{1}{2^k}$. This solves Exercise 3.27 **(b)**. □

7.36. Solution to Exercise 4.1

Solution to Exercise 4.1. We claim that every $n\in\mathbb{N}$ satisfies

$$x_n=\frac{1}{2^n}\left(2na^{n-1}x_1-(n-1)a^nx_0\right) \qquad (695)$$

(where na^{n-1} is to be understood as 0 when $n=0$ [339]).

[*Proof of (695):* We shall prove (695) by strong induction[340] on n. So we fix some $N\in\mathbb{N}$, and we assume that (695) is already proven for every $n<N$. (This is our induction hypothesis.) We now need to show that (695) holds for $n=N$ as well.[341]

[339] This needs to be said, because a^{n-1} alone can be undefined for $n=0$ (if $a=0$).

[340] See Section 2.8 for an introduction to strong induction.

[341] If you are wondering "where is the induction base?": It isn't missing. A strong induction needs no induction base (see Convention 2.63 for the details). Strong induction lets you prove that some statement $\mathcal{A}_n$ holds for every $n\in\mathbb{N}$ by means of proving that for every $N\in\mathbb{N}$,

$$\text{if } \mathcal{A}_n \text{ holds for every } n<N\text{, then } \mathcal{A}_N \text{ holds.} \qquad (696)$$

This immediately shows that $\mathcal{A}_0$ holds: Namely, it is clear that $\mathcal{A}_n$ holds for every $n<0$ (because

In other words, we need to prove that

$$x_N = \frac{1}{2^N} \left(2Na^{N-1}x_1 - (N-1) a^N x_0 \right). \tag{697}$$

We must be in one of the following three cases:

Case 1: We have $N = 0$.

Case 2: We have $N = 1$.

Case 3: We have $N \geq 2$.

Let us first consider Case 1. In this case, we have $N = 0$. Hence,

$$\begin{aligned} \frac{1}{2^N} \left(2Na^{N-1}x_1 - (N-1) a^N x_0 \right) &= \underbrace{\frac{1}{2^0}}_{=1} \left(\underbrace{2 \cdot 0 a^{0-1} x_1}_{=0} - \underbrace{(0-1) a^0}_{=-1} x_0 \right) \\ &= 1 (0 - (-1) x_0) = 1 x_0 = x_0 \\ &= x_N \qquad (\text{since } 0 = N). \end{aligned}$$

In other words, $x_N = \frac{1}{2^N} \left(2Na^{N-1}x_1 - (N-1) a^N x_0 \right)$. Hence, (697) is proven in Case 1.

The proof of (697) in Case 2 is similarly straightforward, and is left to the reader.

Let us now consider Case 3. In this case, we have $N \geq 2$. Hence, both $N-1$ and $N-2$ are nonnegative integers. Moreover, $N - 1 < N$, so that (695) is already proven for $n = N - 1$ (by our induction hypothesis). In other words, we have

$$x_{N-1} = \frac{1}{2^{N-1}} \left(2 (N-1) a^{N-2} x_1 - (N-2) a^{N-1} x_0 \right). \tag{698}$$

Also, $N-2$ is a nonnegative integer such that $N - 2 < N$. Hence, (695) is already proven for $n = N - 2$ (by our induction hypothesis). In other words, we have

$$x_{N-2} = \frac{1}{2^{N-2}} \left(2 (N-2) a^{N-3} x_1 - (N-3) a^{N-2} x_0 \right). \tag{699}$$

Now, recall that $a^2 + 4b = 0$, so that $4b = -a^2$. Hence, $4b \cdot (N-2) a^{N-3} = \left(-a^2\right) \cdot (N-2) a^{N-3} = -(N-2) a^{N-1}$. (Don't forget to check that this latter equality holds also when $N - 2 = 0$; keep in mind that $(N-2) a^{N-3}$ was defined to be 0 in this case, although a^{N-3} might be undefined.) From $4b = -a^2$, we also deduce $b = -\frac{a^2}{4}$, so that $b (N-3) a^{N-2} = \frac{-a^2}{4} (N-3) a^{N-2} = -\frac{1}{4} (N-3) a^N$.

there exists no $n < 0$), and thus (696) (applied to $N = 0$) shows that $\mathcal{A}_0$ holds.

Of course, the proof of (696) might involve some case analysis; in particular, it might argue differently depending on whether $N = 0$ or $N \geq 1$. (Indeed, our proof will be something like this: it will treat the cases $N = 0$, $N = 1$ and $N \geq 2$ separately.) So there can be a "de-facto induction base" (or two, or many) hidden in the proof of (696).

But the sequence $(x_0, x_1, x_2, \ldots)$ is (a,b)-recurrent. Hence,

$$\begin{aligned}
x_N &= ax_{N-1} + bx_{N-2} \\
&= a \cdot \frac{1}{2^{N-1}} \left(2(N-1)a^{N-2}x_1 - (N-2)a^{N-1}x_0\right) \\
&\qquad + b \cdot \frac{1}{2^{N-2}} \left(2(N-2)a^{N-3}x_1 - (N-3)a^{N-2}x_0\right) \\
&\qquad \text{(by (698) and (699))} \\
&= a \cdot \frac{1}{2^{N-1}} 2(N-1)a^{N-2}x_1 - a \cdot \frac{1}{2^{N-1}}(N-2)a^{N-1}x_0 \\
&\qquad + b \cdot \frac{1}{2^{N-2}} 2(N-2)a^{N-3}x_1 - b \cdot \frac{1}{2^{N-2}}(N-3)a^{N-2}x_0 \\
&= \underbrace{\left(a \cdot \frac{1}{2^{N-1}} 2(N-1)a^{N-2} + b \cdot \frac{1}{2^{N-2}} 2(N-2)a^{N-3}\right)}_{=\frac{1}{2^{N-1}}(a \cdot 2(N-1)a^{N-2} + 4b \cdot (N-2)a^{N-3})} x_1 \\
&\qquad - \underbrace{\left(a \cdot \frac{1}{2^{N-1}}(N-2)a^{N-1} + b \cdot \frac{1}{2^{N-2}}(N-3)a^{N-2}\right)}_{=\frac{1}{2^{N-1}}\left(a \cdot (N-2)a^{N-1} + 2b(N-3)a^{N-2}\right)} x_0 \\
&= \frac{1}{2^{N-1}} \left(\underbrace{a \cdot 2(N-1)a^{N-2}}_{=2(N-1)a^{N-1}} + \underbrace{4b \cdot (N-2)a^{N-3}}_{=-(N-2)a^{N-1}}\right) x_1 \\
&\qquad - \frac{1}{2^{N-1}} \left(\underbrace{a \cdot (N-2)a^{N-1}}_{=(N-2)a^N} + 2\underbrace{b(N-3)a^{N-2}}_{=-\frac{1}{4}(N-3)a^N}\right) x_0 \\
&= \frac{1}{2^{N-1}} \underbrace{\left(2(N-1)a^{N-1} + \left(-(N-2)a^{N-1}\right)\right)}_{=Na^{N-1}} x_1 \\
&\qquad - \frac{1}{2^{N-1}} \underbrace{\left((N-2)a^N + 2\left(-\frac{1}{4}(N-3)\right)a^N\right)}_{=\frac{1}{2}(N-1)a^N} x_0 \\
&= \frac{1}{2^{N-1}} Na^{N-1}x_1 - \frac{1}{2^{N-1}} \cdot \frac{1}{2}(N-1)a^N x_0 = \frac{1}{2^{N-1}} \left(Na^{N-1}x_1 - \frac{1}{2}(N-1)a^N x_0\right) \\
&= \frac{1}{2^N} \left(2Na^{N-1}x_1 - (N-1)a^N x_0\right).
\end{aligned}$$

In other words, (697) is proven in Case 3.

Thus we have seen that (697) holds in each of the three Cases 1, 2 and 3. Since these three cases cover all possibilities, this finishes the proof of (697). Hence, we have finished our proof of (695) by strong induction.] □

Remark 7.67. Proving the identity (695) by strong induction is a completely straightforward task. The main difficulty of the exercise is finding this identity. Linear algebra (specifically, the theory of the Jordan normal form) gives a "conceptual" way to derive it (see, e.g., [Fische01] for a very general treatment), but it can also be experimentally found by computing x_2, x_3, x_4, x_5, x_6 directly (using $a^2 + 4b = 0$ to rewrite b as $-\frac{a^2}{4}$, so that only the variable a appears in the expressions) and guessing the pattern.

7.37. Solution to Exercise 4.2

Solution to Exercise 4.2. We shall only solve part **(c)**, since the other two parts are its particular cases (for $N = 2$ and for $N = 3$, respectively).

(c) We define a new sequence $(c_0, c_1, c_2, \ldots)$ recursively by

$$c_0 = 2, \qquad c_1 = a, \qquad \text{and} \qquad c_n = ac_{n-1} + bc_{n-2} \qquad \text{for all } n \geq 2.$$

(So this sequence $(c_0, c_1, c_2, \ldots)$ is (a, b)-recurrent. Its first values are $c_0 = 2$, $c_1 = a$, $c_2 = a^2 + 2b$, $c_3 = a\left(a^2 + 3b\right)$ and $c_4 = a^4 + 4a^2b + 2b^2$. We notice that this sequence depends only on a and b.)

We now claim that every $N \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy

$$x_{m+2N} = c_N x_{m+N} + (-1)^{N-1} b^N x_m. \qquad (700)$$

Once this is proven, we will be done: In fact, (700) shows that, for every nonnegative integers N and K, the sequence $(x_K, x_{N+K}, x_{2N+K}, x_{3N+K}, \ldots)$ is $\left(c_N, (-1)^{N-1} b^N\right)$-recurrent[342]. Thus, in order to solve Exercise 4.2 **(c)**, we only need to prove (700).

[342] *Proof.* Assume that we have already proven (700). Now, for every nonnegative integers N and K, for every $u \geq 2$, we have

$$\begin{aligned} x_{uN+K} &= x_{(u-2)N+K+2N} && (\text{since } uN+K = (u-2)N + K + 2N) \\ &= c_N \underbrace{x_{(u-2)N+K+N}}_{=x_{(u-1)N+K}} + (-1)^{N-1} b^N x_{(u-2)N+K} && (\text{by (700), applied to } m = (u-2)N+K) \\ &= c_N x_{(u-1)N+K} + (-1)^{N-1} b^N x_{(u-2)N+K}. \end{aligned}$$

In other words, for every nonnegative integers N and K, the sequence $(x_K, x_{N+K}, x_{2N+K}, x_{3N+K}, \ldots)$ is $\left(c_N, (-1)^{N-1} b^N\right)$-recurrent. Qed.

[*Proof of (700):* We shall prove (700) by strong induction over N. Thus, we fix some $n \in \mathbb{N}$, and we assume (as our induction hypothesis) that (700) holds for every $N < n$ (and, of course, every $m \in \mathbb{N}$). We need to prove that (700) holds for $N = n$ (and every $m \in \mathbb{N}$). In other words, we need to prove that

$$x_{m+2n} = c_n x_{m+n} + (-1)^{n-1} b^n x_m \tag{701}$$

for every $m \in \mathbb{N}$.

We must be in one of the following three cases:

Case 1: We have $n = 0$.

Case 2: We have $n = 1$.

Case 3: We have $n \geq 2$.

Let us first consider Case 1. In this case, we have $n = 0$. Now, let $m \in \mathbb{N}$. Since $n = 0$, we have

$$c_n x_{m+n} + (-1)^{n-1} b^n x_m = \underbrace{c_0}_{=2} \underbrace{x_{m+0}}_{=x_m} + \underbrace{(-1)^{0-1}}_{=-1} \underbrace{b^0}_{=1} x_m = 2x_m + (-1) x_m = x_m.$$

Compared with

$$\begin{aligned} x_{m+2n} &= x_{m+2\cdot 0} \qquad (\text{since } n = 0) \\ &= x_m, \end{aligned}$$

this yields $x_{m+2n} = c_n x_{m+n} + (-1)^{n-1} b^n x_m$. Hence, (701) is proven in Case 1.

Let us next consider Case 2. In this case, we have $n = 1$. Now, let $m \in \mathbb{N}$. Since $n = 1$, we have

$$c_n x_{m+n} + (-1)^{n-1} b^n x_m = \underbrace{c_1}_{=a} x_{m+1} + \underbrace{(-1)^{1-1}}_{=1} \underbrace{b^1}_{=b} x_m = a x_{m+1} + b x_m.$$

Compared with

$$\begin{aligned} x_{m+2n} &= x_{m+2\cdot 1} \qquad (\text{since } n = 1) \\ &= x_{m+2} = a x_{(m+2)-1} + b x_{(m+2)-2} \\ &\qquad (\text{since the sequence } (x_0, x_1, x_2, \ldots) \text{ is } (a,b)\text{-recurrent}) \\ &= a x_{m+1} + b x_m, \end{aligned}$$

this yields $x_{m+2n} = c_n x_{m+n} + (-1)^{n-1} b^n x_m$. Hence, (701) is proven in Case 2.

Let us finally consider Case 3. In this case, we have $n \geq 2$. Hence, both $n - 1$ and $n - 2$ are nonnegative integers. Moreover, $n - 1 < n$, so that (700) is already proven for $N = n - 1$ (by our induction hypothesis). In other words, we have

$$x_{m+2(n-1)} = c_{n-1} x_{m+(n-1)} + (-1)^{(n-1)-1} b^{n-1} x_m \tag{702}$$

for every $m \in \mathbb{N}$.

Also, $n-2$ is a nonnegative integer such that $n-2<n$. Hence, (695) is already proven for $N=n-2$ (by our induction hypothesis). In other words, we have

$$x_{m+2(n-2)} = c_{n-2} x_{m+(n-2)} + (-1)^{(n-2)-1} b^{n-2} x_m \tag{703}$$

for every $m \in \mathbb{N}$.

Now, fix $m \in \mathbb{N}$. We want to prove (701). The recursive definition of the sequence $(c_0, c_1, c_2, \ldots)$ yields

$$c_n = ac_{n-1} + bc_{n-2}. \tag{704}$$

Since the sequence $(x_0, x_1, x_2, \ldots)$ is (a,b)-recurrent, we have

$$x_{m+2} = a \underbrace{x_{(m+2)-1}}_{=x_{m+1}} + b \underbrace{x_{(m+2)-2}}_{=x_m} = ax_{m+1} + bx_m,$$

so that

$$ax_{m+1} - x_{m+2} = -bx_m. \tag{705}$$

But since the sequence $(x_0, x_1, x_2, \ldots)$ is (a,b)-recurrent, we also have

$$\begin{aligned}
x_{m+2n} &= a \underbrace{x_{(m+2n)-1}}_{\substack{=x_{(m+1)+2(n-1)} \\ (\text{since } (m+2n)-1=(m+1)+2(n-1))}} + b \underbrace{x_{(m+2n)-2}}_{\substack{=x_{(m+2)+2(n-2)} \\ (\text{since } (m+2n)-2=(m+2)+2(n-2))}} \\
&\qquad \left(\text{since } \underbrace{m}_{\geq 0} + 2 \underbrace{n}_{\geq 2} \geq 0 + 2 \cdot 2 = 4 \geq 2\right) \\
&= a \underbrace{x_{(m+1)+2(n-1)}}_{\substack{=c_{n-1}x_{(m+1)+(n-1)}+(-1)^{(n-1)-1}b^{n-1}x_{m+1} \\ (\text{by (702), applied to } m+1 \text{ instead of } m)}} + b \underbrace{x_{(m+2)+2(n-2)}}_{\substack{=c_{n-2}x_{(m+2)+(n-2)}+(-1)^{(n-2)-1}b^{n-2}x_{m+2} \\ (\text{by (703), applied to } m+2 \text{ instead of } m)}} \\
&= a \left(c_{n-1} \underbrace{x_{(m+1)+(n-1)}}_{=x_{m+n}} + \underbrace{(-1)^{(n-1)-1}}_{=(-1)^{n-2}=(-1)^n} b^{n-1} x_{m+1} \right) \\
&\qquad + b \left(c_{n-2} \underbrace{x_{(m+2)+(n-2)}}_{=x_{m+n}} + \underbrace{(-1)^{(n-2)-1}}_{=(-1)^{n-3}=(-1)^{n-1}=-(-1)^n} b^{n-2} x_{m+2} \right) \\
&= a \left(c_{n-1}x_{m+n} + (-1)^n b^{n-1} x_{m+1} \right) + b \left(c_{n-2}x_{m+n} + \left(-(-1)^n\right) b^{n-2} x_{m+2} \right) \\
&= ac_{n-1}x_{m+n} + (-1)^n ab^{n-1}x_{m+1} + bc_{n-2}x_{m+n} - (-1)^n \underbrace{bb^{n-2}}_{=b^{n-1}} x_{m+2} \\
&= ac_{n-1}x_{m+n} + (-1)^n ab^{n-1}x_{m+1} + bc_{n-2}x_{m+n} - (-1)^n b^{n-1} x_{m+2} \\
&= \underbrace{\left(ac_{n-1}x_{m+n} + bc_{n-2}x_{m+n}\right)}_{=(ac_{n-1}+bc_{n-2})x_{m+n}} + \underbrace{(-1)^n ab^{n-1}x_{m+1} - (-1)^n b^{n-1} x_{m+2}}_{=(-1)^n b^{n-1}(ax_{m+1}-x_{m+2})} \\
&= \underbrace{\left(ac_{n-1} + bc_{n-2}\right)}_{\substack{=c_n \\ (\text{by (704)})}} x_{m+n} + (-1)^n b^{n-1} \underbrace{\left(ax_{m+1} - x_{m+2}\right)}_{\substack{=-bx_m \\ (\text{by (705)})}} \\
&= c_n x_{m+n} + \underbrace{(-1)^n b^{n-1} \left(-bx_m\right)}_{=-(-1)^n b^{n-1} bx_m} = c_n x_{m+n} + \left(\underbrace{-(-1)^n}_{=(-1)^{n-1}} \underbrace{b^{n-1}b}_{=b^n} x_m \right) \\
&= c_n x_{m+n} + (-1)^{n-1} b^n x_m .
\end{aligned}$$

In other words, (701) is proven in Case 3.

Thus we have seen that (701) holds in each of the three Cases 1, 2 and 3. Since these three cases cover all possibilities, this finishes the proof of (701). This finishes our (inductive) proof of (700).]

As we know, this solves Exercise 4.2. □

> **Remark 7.68.** How on earth could one have come up with my definition of the sequence $(c_0, c_1, c_2, \ldots)$ in the solution above? One way is to solve parts **(a)** and **(b)** of the exercise first (which can be solved by applying the equation $x_n = ax_{n-1} + bx_{n-2}$ several times), and then guess that the answer to **(c)** is a pair of the form $(c, d) = \left(c_N, (-1)^{N-1} b^N\right)$ for some sequence $(c_0, c_1, c_2, \ldots)$. What remains is finding this sequence. I believe its entry $c_3 = a\left(a^2 + 3b\right)$ is particularly telltale.

7.38. Solution to Exercise 4.3

Solution to Exercise 4.3. A set I of integers is said to be *lacunar* if no two elements of I are consecutive (i.e., there exists no $i \in \mathbb{Z}$ such that both i and $i+1$ belong to I). Then, Exercise 4.3 asks us to prove that, for every positive integer n,

$$\text{the number } f_n \text{ is the number of lacunar subsets of } \{1, 2, \ldots, n-2\}. \qquad (706)$$

We shall prove (706) by strong induction over n. Thus, we let N be a positive integer, and we assume (as the induction hypothesis) that (706) is proven for every $n < N$. We need to prove (706) for $n = N$. In other words, we need to prove that

$$\text{the number } f_N \text{ is the number of lacunar subsets of } \{1, 2, \ldots, N-2\}. \qquad (707)$$

Recall that N is a positive integer. Hence, we are in one of the following three cases:

Case 1: We have $N = 1$.

Case 2: We have $N = 2$.

Case 3: We have $N \geq 3$.

Let us first consider Case 1. In this case, we have $N = 1$. Thus, $f_N = f_1 = 1$. On the other hand, the number of lacunar subsets of $\{1, 2, \ldots, N-2\}$ is 1 (since the set $\left\{1, 2, \ldots, \underbrace{N}_{=1} - 2\right\} = \{1, 2, \ldots, 1-2\} = \varnothing$ has only one subset, and this subset is lacunar). Thus, f_N is the number of lacunar subsets of $\{1, 2, \ldots, N-2\}$. Hence, (707) is proven in Case 1.

Case 2 can be dealt with similarly (in this case, the set $\{1, 2, \ldots, N-2\}$ is still empty, and f_N is still 1), and is left to the reader.

We now consider Case 3. In this case, we have $N \geq 3$. Hence, $N-1$ and $N-2$ are positive integers. Since $N-1$ is a positive integer and $< N$, we know that (707) is proven for $n = N-1$ (due to our induction hypothesis). In other words, the number f_{N-1} is the number of lacunar subsets of $\{1, 2, \ldots, (N-1)-2\}$. In other words, f_{N-1} is the number of lacunar subsets of $\{1, 2, \ldots, N-3\}$.

Also, since $N-2$ is a positive integer and $< N$, we know that (707) is proven for $n = N-2$ (due to our induction hypothesis). In other words, the number f_{N-2} is the number of lacunar subsets of $\{1, 2, \ldots, (N-2) - 2\}$. In other words, f_{N-2} is the number of lacunar subsets of $\{1, 2, \ldots, N-4\}$.

Now, how do the lacunar subsets of $\{1, 2, \ldots, N-2\}$ look like? We say that a lacunar subset of $\{1, 2, \ldots, N-2\}$ has *type 1* if it contains $N-2$, and has *type 2* if it does not. Now, let us count the lacunar subsets having type 1 and those having type 2:

- If I is a lacunar subset of $\{1, 2, \ldots, N-2\}$ which has type 1, then it contains $N-2$, and thus cannot contain $N-3$ (because it is lacunar, i.e., contains no two consecutive integers, but $N-3$ and $N-2$ are two consecutive integers); moreover, $I \setminus \{N-2\}$ is a lacunar subset of $\{1, 2, \ldots, N-4\}$ [343]. Thus, to every lacunar subset I of $\{1, 2, \ldots, N-2\}$ which has type 1, we have assigned a lacunar subset of $\{1, 2, \ldots, N-4\}$ (namely, $I \setminus \{N-2\}$). It is easy to see that this assignment is injective (indeed, if I and J are two lacunar subsets of $\{1, 2, \ldots, N-2\}$ which have type 1, and if $I \setminus \{N-2\} = J \setminus \{N-2\}$, then $I = J$) and surjective (because whenever K is a lacunar subset of $\{1, 2, \ldots, N-4\}$, the set $K \cup \{N-2\}$ is a lacunar subset of $\{1, 2, \ldots, N-2\}$ which has type 1, and this set $K \cup \{N-2\}$ is sent back to K by our assignment); thus, it is bijective. Hence, we have found a bijection between the lacunar subsets of $\{1, 2, \ldots, N-2\}$ which have type 1 and the lacunar subsets of $\{1, 2, \ldots, N-4\}$. Therefore, the number of lacunar subsets of $\{1, 2, \ldots, N-2\}$ which have type 1 equals the number of lacunar subsets of $\{1, 2, \ldots, N-4\}$. But the latter number is f_{N-2}. Hence, the number of lacunar subsets of $\{1, 2, \ldots, N-2\}$ which have type 1 is f_{N-2}.

- The lacunar subsets of $\{1, 2, \ldots, N-2\}$ which have type 2 are precisely the lacunar subsets of $\{1, 2, \ldots, N-2\}$ which do not contain $N-2$; in other words, they are precisely the lacunar subsets of $\{1, 2, \ldots, N-3\}$. Hence, the number of lacunar subsets of $\{1, 2, \ldots, N-2\}$ which have type 2 is f_{N-1} (since we know that the number of lacunar subsets of $\{1, 2, \ldots, N-3\}$ is f_{N-1}).

Now, let us summarize. Each of the lacunar subsets of $\{1, 2, \ldots, N-2\}$ either has type 1 or has type 2 (but not both). Hence, the number of lacunar subsets of $\{1, 2, \ldots, N-2\}$ equals

$$\begin{aligned}
&\underbrace{(\text{the number of lacunar subsets of } \{1, 2, \ldots, N-2\} \text{ which have type 1})}_{=f_{N-2}} \\
&\qquad + \underbrace{(\text{the number of lacunar subsets of } \{1, 2, \ldots, N-2\} \text{ which have type 2})}_{=f_{N-1}} \\
&= f_{N-2} + f_{N-1} = f_{N-1} + f_{N-2} \\
&= f_N \qquad (\text{by the recursion of the Fibonacci numbers}).
\end{aligned}$$

[343] Indeed, it is lacunar because I is lacunar; and it is a subset of $\{1, 2, \ldots, N-4\}$ because I cannot contain $N-3$.

This proves (707) in Case 3.

Now, (707) is proven in each of the three Cases 1, 2 and 3. Hence, (707) always holds. This completes our (inductive) proof of (706). Exercise 4.3 is solved. □

7.39. Solution to Exercise 4.4

7.39.1. The solution

Solution to Exercise 4.4. Let us first explain why the right hand side of (291) is well-defined. In fact, this is not obvious, because if $r = 0$, then r^{n-1-2i} might not always make sense (because $n-1-2i$ can be negative). However, it turns out that $\binom{n-1-i}{i} = 0$ for every $i \in \{0, 1, \ldots, n-1\}$ satisfying $n - 1 - 2i < 0$ [344]. Hence, we interpret $\binom{n-1-i}{i} r^{n-1-2i}$ as 0 whenever $n-1-2i < 0$ (even if the term r^{n-1-2i} by itself is not well-defined), following our convention that any expression of the form $a \cdot b$ with $a = 0$ has to be interpreted as 0. Thus, the right hand side of (291) is well-defined.

Now, we shall prove (291) by strong induction over n. Thus, we let $N \in \mathbb{N}$, and we assume (as the induction hypothesis) that (291) is proven for every $n < N$. We need to prove (291) for $n = N$. In other words, we need to prove that

$$c_N = \sum_{i=0}^{N-1} (-1)^i \binom{N-1-i}{i} r^{N-1-2i}. \tag{708}$$

Recall that $N \in \mathbb{N}$. Hence, we are in one of the following three cases:

Case 1: We have $N = 0$.

Case 2: We have $N = 1$.

Case 3: We have $N \geq 2$.

Let us first consider Case 1. In this case, we have $N = 0$. Hence, $c_N = c_0 = 0$. But also, the sum $\sum_{i=0}^{N-1} (-1)^i \binom{N-1-i}{i} r^{N-1-2i}$ is an empty sum (since $N = 0$) and thus equals 0. Therefore, both sides of the equality (708) equal 0. Hence, the equality (708) holds. We thus have proven (708) in Case 1.

Let us now consider Case 2. In this case, we have $N = 1$. Thus, $c_N = c_1 = 1$. On

[344] *Proof.* Let $i \in \{0, 1, \ldots, n-1\}$ be such that $n - 1 - 2i < 0$. Then, $i \leq n-1$ (since $i \in \{0, 1, \ldots, n-1\}$), so that $n - 1 - i \geq 0$. In other words, $n-1-i \in \mathbb{N}$. But $n - 1 - i < i$ (since $(n-1-i) - i = n - 1 - 2i < 0$). Hence, Proposition 3.6 (applied to $n-1-i$ and i instead of m and n) yields $\binom{n-1-i}{i} = 0$. Qed.

the other hand, from $N = 1$, we obtain

$$\sum_{i=0}^{N-1} (-1)^i \binom{N-1-i}{i} r^{N-1-2i} = \sum_{i=0}^{1-1} (-1)^i \binom{1-1-i}{i} r^{1-1-2i}$$
$$= \underbrace{(-1)^0}_{=1} \underbrace{\binom{1-1-0}{0}}_{=1} \underbrace{r^{1-1-2\cdot 0}}_{=r^0=1} = 1 = c_N.$$

Hence, (708) is proven in Case 2.

Let us now consider Case 3. In this case, we have $N \geq 2$. Therefore, both $N-1$ and $N-2$ belong to $\mathbb{N}$.

So we know that $N-1$ is an element of $\mathbb{N}$ satisfying $N-1 < N$. Hence, (291) is proven for $n = N-1$ (by our induction hypothesis). In other words,

$$c_{N-1} = \sum_{i=0}^{N-2} (-1)^i \binom{N-2-i}{i} r^{N-2-2i}. \qquad (709)$$

Also, $N-2$ is an element of $\mathbb{N}$ satisfying $N-2 < N$. Hence, (291) is proven for $n = N-2$ (by our induction hypothesis). In other words,

$$c_{N-2} = \sum_{i=0}^{N-3} (-1)^i \binom{N-3-i}{i} r^{N-3-2i}. \qquad (710)$$

Let us further recall that every positive integer i and every $a \in \mathbb{Z}$ satisfy

$$\binom{a-1}{i} + \binom{a-1}{i-1} = \binom{a}{i} \qquad (711)$$

[345].

[345] *Proof of (711):* Let i be a positive integer. Let $a \in \mathbb{Z}$. Then, (234) (applied to $m = a$ and $n = i$) shows that $\binom{a}{i} = \binom{a-1}{i-1} + \binom{a-1}{i} = \binom{a-1}{i} + \binom{a-1}{i-1}$. This proves (711).

Now, $N \geq 2$, so that the recursive definition of the sequence $(c_0, c_1, c_2, \ldots)$ yields

$$c_N = r \underbrace{c_{N-1}}_{\substack{=\sum_{i=0}^{N-2} (-1)^i \binom{N-2-i}{i} r^{N-2-2i} \\ \text{(by (709))}}} - \underbrace{c_{N-2}}_{\substack{=\sum_{i=0}^{N-3} (-1)^i \binom{N-3-i}{i} r^{N-3-2i} \\ \text{(by (710))}}}$$

$$= r \left(\sum_{i=0}^{N-2} (-1)^i \binom{N-2-i}{i} r^{N-2-2i} \right) - \sum_{i=0}^{N-3} (-1)^i \binom{N-3-i}{i} r^{N-3-2i}$$

$$= \sum_{i=0}^{N-2} (-1)^i \binom{N-2-i}{i} \underbrace{r r^{N-2-2i}}_{=r^{N-1-2i}} - \underbrace{\sum_{i=0}^{N-3} (-1)^i \binom{N-3-i}{i} r^{N-3-2i}}_{\substack{=\sum_{i=1}^{N-2} (-1)^{i-1} \binom{N-3-(i-1)}{i-1} r^{N-3-2(i-1)} \\ \text{(here, we substituted } i-1 \text{ for } i \text{ in the sum)}}}$$

$$= \underbrace{\sum_{i=0}^{N-2} (-1)^i \binom{N-2-i}{i} r^{N-1-2i}}_{=(-1)^0 \binom{N-2-0}{0} r^{N-1-2\cdot 0} + \sum_{i=1}^{N-2} (-1)^i \binom{N-2-i}{i} r^{N-1-2i}}$$

$$- \sum_{i=1}^{N-2} \underbrace{(-1)^{i-1}}_{=-(-1)^i} \underbrace{\binom{N-3-(i-1)}{i-1}}_{=\binom{N-2-i}{i-1}} \underbrace{r^{N-3-2(i-1)}}_{=r^{N-1-2i}}$$

$$= (-1)^0 \underbrace{\binom{N-2-0}{0}}_{=1=\binom{N-1-0}{0}} r^{N-1-2\cdot 0}$$
$$+ \underbrace{\sum_{i=1}^{N-2} (-1)^i \binom{N-2-i}{i} r^{N-1-2i} - \sum_{i=1}^{N-2} \left(-(-1)^i\right) \binom{N-2-i}{i-1} r^{N-1-2i}}_{=\sum_{i=1}^{N-2} (-1)^i \left(\binom{N-2-i}{i} + \binom{N-2-i}{i-1}\right) r^{N-1-2i}}$$
$$= (-1)^0 \binom{N-1-0}{0} r^{N-1-2\cdot 0}$$
$$+ \sum_{i=1}^{N-2} (-1)^i \underbrace{\left(\binom{N-2-i}{i} + \binom{N-2-i}{i-1}\right)}_{\substack{=\binom{(N-1-i)-1}{i}+\binom{(N-1-i)-1}{i-1} \\ =\binom{N-1-i}{i} \\ \text{(by (711), applied to } a=N-1-i)}} r^{N-1-2i}$$
$$= (-1)^0 \binom{N-1-0}{0} r^{N-1-2\cdot 0} + \sum_{i=1}^{N-2} (-1)^i \binom{N-1-i}{i} r^{N-1-2i}$$
$$= \sum_{i=0}^{N-2} (-1)^i \binom{N-1-i}{i} r^{N-1-2i}.$$

Hence, (708) is proven in Case 3. We thus have proven (708) in all three Cases 1, 2 and 3, so that we conclude that (708) always holds. This completes our proof of (291). ☐

7.39.2. A corollary

Having solved Exercise 4.4, let us show an identity for binomial coefficients that follows from it:

Corollary 7.69. For each $n \in \mathbb{N}$, we have

$$\sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} = (-1)^n \cdot \begin{cases} 0, & \text{if } n \equiv 0 \bmod 3; \\ -1, & \text{if } n \equiv 1 \bmod 3; \\ 1, & \text{if } n \equiv 2 \bmod 3 \end{cases}.$$

Proof of Corollary 7.69. Define $r \in \mathbb{Z}$ by $r = -1$. Define a sequence $(c_0, c_1, c_2, \ldots)$ of integers as in Exercise 4.4. (Thus, $c_0 = 0$, $c_1 = 1$ and $c_n = rc_{n-1} - c_{n-2}$ for all $n \geq 2$.)

Observation 1: We have

$$c_n = \begin{cases} 0, & \text{if } n \equiv 0 \bmod 3; \\ 1, & \text{if } n \equiv 1 \bmod 3; \\ -1, & \text{if } n \equiv 2 \bmod 3 \end{cases}$$

for each $n \in \mathbb{N}$.

[*Proof of Observation 1:* We shall prove Observation 1 by strong induction over n:

Induction step: Let $N \in \mathbb{N}$. Assume that Observation 1 holds whenever $n < N$. We must now prove that Observation 1 holds for $n = N$. In other words, we must prove that

$$c_N = \begin{cases} 0, & \text{if } N \equiv 0 \bmod 3; \\ 1, & \text{if } N \equiv 1 \bmod 3; \\ -1, & \text{if } N \equiv 2 \bmod 3 \end{cases}. \tag{712}$$

If $N < 2$, then this can be checked directly (since $c_0 = 0$ and $c_1 = 1$). Thus, for the rest of this proof, we WLOG assume that we don't have $N < 2$. Hence, $N \geq 2$; therefore, $N - 1 \in \mathbb{N}$ and $N - 2 \in \mathbb{N}$.

But the recursive definition of the sequence $(c_0, c_1, c_2, \ldots)$ yields $c_n = rc_{n-1} - c_{n-2}$ for all $n \geq 2$. Applying this to $n = N$, we find

$$c_N = \underbrace{r}_{=-1} c_{N-1} - c_{N-2} = (-1) c_{N-1} - c_{N-2} = -c_{N-1} - c_{N-2}. \tag{713}$$

We have assumed that Observation 1 holds whenever $n < N$. Thus, we can apply Observation 1 to $n = N - 1$ (since $N - 1 < N$ and $N - 1 \in \mathbb{N}$). We thus conclude that

$$c_{N-1} = \begin{cases} 0, & \text{if } N-1 \equiv 0 \bmod 3; \\ 1, & \text{if } N-1 \equiv 1 \bmod 3; \\ -1, & \text{if } N-1 \equiv 2 \bmod 3 \end{cases}. \tag{714}$$

Similarly, we obtain

$$c_{N-2} = \begin{cases} 0, & \text{if } N-2 \equiv 0 \bmod 3; \\ 1, & \text{if } N-2 \equiv 1 \bmod 3; \\ -1, & \text{if } N-2 \equiv 2 \bmod 3 \end{cases}. \tag{715}$$

The remainder obtained when N is divided by 3 must be either 0 or 1 or 2. Hence, we must have either $N \equiv 0 \bmod 3$ or $N \equiv 1 \bmod 3$ or $N \equiv 2 \bmod 3$. Thus, we are in one of the following three cases:

Case 1: We have $N \equiv 0 \bmod 3$.

Case 2: We have $N \equiv 1 \bmod 3$.

Case 3: We have $N \equiv 2 \bmod 3$.

We shall only consider Case 1 (and leave the two other cases, which are analogous, to the reader). In this case, we have $N \equiv 0 \mod 3$. Hence, $\underbrace{N}_{\equiv 0 \mod 3} -1 \equiv 0 - 1 \equiv 2 \mod 3$. Thus, (714) simplifies to $c_{N-1} = -1$. Also, $\underbrace{N}_{\equiv 0 \mod 3} -2 \equiv 0 - 2 \equiv 1 \mod 3$. Thus, (715) simplifies to $c_{N-2} = 1$. Now, (713) yields

$$c_N = -\underbrace{c_{N-1}}_{=-1} - \underbrace{c_{N-2}}_{=1} = -(-1) - 1 = 0.$$

Comparing this with

$$\begin{cases} 0, & \text{if } N \equiv 0 \mod 3; \\ 1, & \text{if } N \equiv 1 \mod 3; \\ -1, & \text{if } N \equiv 2 \mod 3 \end{cases} = 0 \qquad (\text{since } N \equiv 0 \mod 3),$$

we obtain $c_N = \begin{cases} 0, & \text{if } N \equiv 0 \mod 3; \\ 1, & \text{if } N \equiv 1 \mod 3; \\ -1, & \text{if } N \equiv 2 \mod 3 \end{cases}$. Thus, (712) is proven in Case 1. As we have said, the proof in the other two cases is analogous. Hence, (712) is proven.

In other words, Observation 1 holds for $n = N$. This completes the induction step. Thus, Observation 1 is proven.]

Now, let $n \in \mathbb{N}$. Clearly, $(-1)^{n-1} (-1)^{n-1} = \left(\underbrace{(-1)(-1)}_{=1}\right)^{n-1} = 1^{n-1} = 1$.

Observation 1 yields

$$c_n = \begin{cases} 0, & \text{if } n \equiv 0 \mod 3; \\ 1, & \text{if } n \equiv 1 \mod 3; \\ -1, & \text{if } n \equiv 2 \mod 3 \end{cases}.$$

Thus,

$$\begin{aligned} -c_n &= -\begin{cases} 0, & \text{if } n \equiv 0 \mod 3; \\ 1, & \text{if } n \equiv 1 \mod 3; \\ -1, & \text{if } n \equiv 2 \mod 3 \end{cases} = \begin{cases} -0, & \text{if } n \equiv 0 \mod 3; \\ -1, & \text{if } n \equiv 1 \mod 3; \\ -(-1), & \text{if } n \equiv 2 \mod 3 \end{cases} \\ &= \begin{cases} 0, & \text{if } n \equiv 0 \mod 3; \\ -1, & \text{if } n \equiv 1 \mod 3; \\ 1, & \text{if } n \equiv 2 \mod 3 \end{cases}. \end{aligned} \tag{716}$$

Exercise 4.4 yields

$$\begin{aligned} c_n &= \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} \underbrace{r^{n-1-2i}}_{\substack{=(-1)^{n-1-2i} \\ \text{(since } r=-1)}} \\ &= \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} \underbrace{(-1)^{n-1-2i}}_{\substack{=(-1)^{n-1} \\ \text{(since } n-1-2i \equiv n-1 \bmod 2)}} \\ &= \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} (-1)^{n-1}. \end{aligned}$$

Multiplying both sides of this equality by $(-1)^{n-1}$, we obtain

$$\begin{aligned} (-1)^{n-1} c_n &= (-1)^{n-1} \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} (-1)^{n-1} \\ &= \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} \underbrace{(-1)^{n-1} (-1)^{n-1}}_{=1} \\ &= \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i}. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=0}^{n-1} (-1)^i \binom{n-1-i}{i} &= \underbrace{(-1)^{n-1}}_{=-(-1)^n} c_n = -(-1)^n c_n \\ &= (-1)^n \cdot \underbrace{(-c_n)}_{\substack{= \begin{cases} 0, & \text{if } n \equiv 0 \bmod 3; \\ -1, & \text{if } n \equiv 1 \bmod 3; \\ 1, & \text{if } n \equiv 2 \bmod 3 \end{cases} \\ \text{(by (716))}}} \\ &= (-1)^n \cdot \begin{cases} 0, & \text{if } n \equiv 0 \bmod 3; \\ -1, & \text{if } n \equiv 1 \bmod 3; \\ 1, & \text{if } n \equiv 2 \bmod 3 \end{cases}. \end{aligned}$$

This proves Corollary 7.69. □

Let us remark that Corollary 7.69 is better known in the following form:

Corollary 7.70. Let $n \in \mathbb{N}$. Then,

$$\sum_{i=0}^{n} (-1)^i \binom{n-i}{i} = \begin{cases} 1, & \text{if } n \equiv 0 \bmod 6 \text{ or } n \equiv 1 \bmod 6; \\ 0, & \text{if } n \equiv 2 \bmod 6 \text{ or } n \equiv 5 \bmod 6; \\ -1, & \text{if } n \equiv 3 \bmod 6 \text{ or } n \equiv 4 \bmod 6 \end{cases}.$$

Proof of Corollary 7.70. Corollary 7.69 (applied to $n+1$ instead of n) yields

$$\sum_{i=0}^{(n+1)-1} (-1)^i \binom{(n+1)-1-i}{i} = (-1)^{n+1} \cdot \begin{cases} 0, & \text{if } n+1 \equiv 0 \bmod 3; \\ -1, & \text{if } n+1 \equiv 1 \bmod 3; \\ 1, & \text{if } n+1 \equiv 2 \bmod 3 \end{cases}.$$

In view of $(n+1)-1 = n$, this rewrites as

$$\sum_{i=0}^{n} (-1)^i \binom{n-i}{i} = (-1)^{n+1} \cdot \begin{cases} 0, & \text{if } n+1 \equiv 0 \bmod 3; \\ -1, & \text{if } n+1 \equiv 1 \bmod 3; \\ 1, & \text{if } n+1 \equiv 2 \bmod 3 \end{cases}. \tag{717}$$

All that remains to be done now is to prove that the right hand side of (717) equals

$$\begin{cases} 1, & \text{if } n \equiv 0 \bmod 6 \text{ or } n \equiv 1 \bmod 6; \\ 0, & \text{if } n \equiv 2 \bmod 6 \text{ or } n \equiv 5 \bmod 6; \\ -1, & \text{if } n \equiv 3 \bmod 6 \text{ or } n \equiv 4 \bmod 6 \end{cases}.$$

This can be verified by straightforward computation, distinguishing six possible cases (one for each remainder n can leave upon division by 6).

Thus, Corollary 7.70 is proven. □

See also [BenQui03, Identity 172] or [BenQui08] for a combinatorial proof of Corollary 7.70. Also, Corollary 7.70 appears in [GrKnPa94, §5.2, Problem 3].

7.40. Solution to Exercise 5.1

Solution to Exercise 5.1. **(a)** This proof is completely straightforward, and would be left to the reader in any research paper; we give a few details only:

Let $i \in \{1, 2, \ldots, n-2\}$. We need to prove that $s_i \circ s_{i+1} \circ s_i = s_{i+1} \circ s_i \circ s_{i+1}$. In order to do so, it is clearly enough to show that $(s_i \circ s_{i+1} \circ s_i)(h) = (s_{i+1} \circ s_i \circ s_{i+1})(h)$ for every $h \in \{1, 2, \ldots, n\}$. So let us fix $h \in \{1, 2, \ldots, n\}$. We must be in one of the following four cases:

Case 1: We have $h = i$.

Case 2: We have $h = i+1$.

Case 3: We have $h = i+2$.

Case 4: We have $h \notin \{i, i+1, i+2\}$.

Let us first consider Case 1. In this case, we have $h = i$ and thus

$$(s_i \circ s_{i+1} \circ s_i)\left(\underbrace{h}_{=i}\right) = (s_i \circ s_{i+1} \circ s_i)(i) = s_i\left(s_{i+1}\left(\underbrace{s_i(i)}_{=i+1}\right)\right)$$
$$= s_i\left(\underbrace{s_{i+1}(i+1)}_{=i+2}\right) = s_i(i+2) = i+2.$$

A similar computation shows $(s_{i+1} \circ s_i \circ s_{i+1})(h) = i+2$. Thus, $(s_i \circ s_{i+1} \circ s_i)(h) = i+2 = (s_{i+1} \circ s_i \circ s_{i+1})(h)$. Hence, we have proven the equality $(s_i \circ s_{i+1} \circ s_i)(h) = (s_{i+1} \circ s_i \circ s_{i+1})(h)$ in Case 1.

Similarly, we can prove the same equality in Cases 2 and 3.

Now, let us consider Case 4. In this case, we have $h \notin \{i, i+1, i+2\}$. Thus, h is neither i nor $i+1$, so that we have $s_i(h) = h$. Also, h is neither $i+1$ nor $i+2$ (since $h \notin \{i, i+1, i+2\}$), and thus we have $s_{i+1}(h) = h$. Hence,

$$(s_i \circ s_{i+1} \circ s_i)(h) = s_i\left(s_{i+1}\left(\underbrace{s_i(h)}_{=h}\right)\right) = s_i\left(\underbrace{s_{i+1}(h)}_{=h}\right) = s_i(h) = h.$$

Similarly, $(s_{i+1} \circ s_i \circ s_{i+1})(h) = h$. Thus, $(s_i \circ s_{i+1} \circ s_i)(h) = h = (s_{i+1} \circ s_i \circ s_{i+1})(h)$. Hence, we have proven $(s_i \circ s_{i+1} \circ s_i)(h) = (s_{i+1} \circ s_i \circ s_{i+1})(h)$ in Case 4.

Altogether, we have proven the equality $(s_i \circ s_{i+1} \circ s_i)(h) = (s_{i+1} \circ s_i \circ s_{i+1})(h)$ in each of the four Cases 1, 2, 3, and 4. Thus, $(s_i \circ s_{i+1} \circ s_i)(h) = (s_{i+1} \circ s_i \circ s_{i+1})(h)$ always holds. Exercise 5.1 **(a)** is thus solved.

(b) We follow the hint and delay the solution of this part until later (see Exercise 5.2 **(e)**).

(c) For every $i \in \{1, 2, \ldots, n\}$, we let a_i be the permutation $s_{i-1} \circ s_{i-2} \circ \cdots \circ s_1 \in S_n$ [346]. Now, we claim that

$$w_0 = a_1 \circ a_2 \circ \cdots \circ a_n. \tag{718}$$

This is essentially an explicit way to write w_0 as a composition of several permutations of the form s_i (because each a_i on the right hand side is the composition $s_{i-1} \circ s_{i-2} \circ \cdots \circ s_1$). Thus, once (718) is proven, the exercise will be solved.

Before we prove (718), let us first understand how a_i operates. We claim that

$$a_i(k) = \begin{cases} i, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq i; \\ k, & \text{if } k > i \end{cases} \tag{719}$$

[346] In particular, $a_1 = s_{1-1} \circ s_{1-2} \circ \cdots \circ s_1$ is the composition of 0 permutations. What does this mean? Just as a sum of 0 terms is defined to be 0 (because 0 is the neutral element of addition), and a product of 0 terms is 1 (since 1 is the neutral element of multiplication), the composition of 0 permutations is defined to be the identity permutation (since the identity permutation is the neutral element of composition). Hence, a_1 is the identity permutation, i.e., we have $a_1 = \mathrm{id}$.

for each $i \in \{1, 2, \ldots, n\}$ and $k \in \{1, 2, \ldots, n\}$. (In other words, a_i is the permutation which sends $1, 2, 3, \ldots, i$ to $i, 1, 2, \ldots, i-1$ and leaves all numbers $> i$ untouched. Using the terminology of Definition 5.37, this means that $a_i = \operatorname{cyc}_{i,i-1,\ldots,1}$.)

[*Proof of (719):* Let me first give an informal proof of (719) which is not hard to turn into a formal proof.

Let $i \in \{1, 2, \ldots, n\}$. We have $a_i = s_{i-1} \circ s_{i-2} \circ \cdots \circ s_1$. Therefore, a_i is the permutation that first switches 1 with 2, then switches 2 with 3, etc., until it finally switches $i-1$ with i. Thus:

- When we apply a_i to 1, we arrive at i at the end (since the 1 is carried to 2 by the first switch, which is then carried to 3 by the next switch, and so on, until it finally becomes i).
- When we apply a_i to some $k \in \{2, 3, \ldots, i\}$, we arrive at $k-1$ at the end (since the first switch to move k is the $(k-1)$-st switch, which changes it into $k-1$, and from then on all the following switches leave $k-1$ untouched).
- When we apply a_i to some $k \in \{i+1, i+2, \ldots, n\}$, we arrive at k at the end (since none of our switches changes k).

Expressing this in a formula instead of in words, we obtain precisely (719).

Formal proof of (719): For the sake of completeness, let me show how to prove (719) formally.

We shall prove (719) by induction on i:

Induction base: We have $a_1 = s_0 \circ s_{-1} \circ \cdots \circ s_1 = (\text{a composition of 0 permutations}) = \operatorname{id}$. Thus, every $k \in \{1, 2, \ldots, n\}$ satisfies

$$\underbrace{a_1}_{=\operatorname{id}}(k) = \operatorname{id}(k) = k = \begin{cases} k, & \text{if } k = 1; \\ k, & \text{if } k > 1 \end{cases} = \begin{cases} 1, & \text{if } k = 1; \\ k, & \text{if } k > 1 \end{cases}$$
$$(\text{since } k = 1 \text{ when } k = 1)$$
$$= \begin{cases} 1, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq 1; \\ k, & \text{if } k > 1 \end{cases}$$
$$\left(\begin{array}{c}\text{here, we added a ``}1 < k \leq 1\text{'' case, which does not} \\ \text{change the result because this case never happens}\end{array}\right).$$

In other words, (719) holds for $i = 1$. This completes the induction base.

Induction step: Let $I \in \{1, 2, \ldots, n\}$ be such that $I > 1$. Assume that (719) holds for $i = I - 1$. We need to show that (719) holds for $i = I$.

We have assumed that (719) holds for $i = I - 1$. In other words,

$$a_{I-1}(k) = \begin{cases} I-1, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I-1; \\ k, & \text{if } k > I-1 \end{cases} \tag{720}$$

for each $k \in \{1, 2, \ldots, n\}$.

The definition of a_{I-1} yields $a_{I-1} = s_{I-2} \circ s_{I-3} \circ \cdots \circ s_1$. The definition of a_I yields $a_I = s_{I-1} \circ s_{I-2} \circ \cdots \circ s_1 = s_{I-1} \circ \underbrace{(s_{I-2} \circ s_{I-3} \circ \cdots \circ s_1)}_{=a_{I-1}} = s_{I-1} \circ a_{I-1}$. Hence, every $k \in \{1, 2, \ldots, n\}$ satisfies

$$a_I(k) = \begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases} \tag{721}$$

[347]. In other words, (719) holds for $i = I$. This completes the induction step. The induction proof of (719) is thus finished.]

---

[347] *Proof of (721):* Let $k \in \{1, 2, \ldots, n\}$. We need to prove (721). We are in one of the following four cases:

Case 1: We have $k = 1$.

Case 2: We have $1 < k \leq I$ and $k < I$.

Case 3: We have $1 < k \leq I$ and $k \geq I$.

Case 4: We have $k > I$.

Let us first consider Case 1. In this case, we have $k = 1$. Thus, (720) yields $a_{I-1}(k) = \begin{cases} I-1, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I-1; \\ k, & \text{if } k > I-1 \end{cases} = I - 1$ (since $k = 1$) and thus $\underbrace{a_I}_{=s_{I-1} \circ a_{I-1}}(k) = (s_{I-1} \circ a_{I-1})(k) = s_{I-1}\left(\underbrace{a_{I-1}(k)}_{=I-1}\right) = s_{I-1}(I-1) = I$. Compared with $\begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases} = I$ (since $k = 1$), this yields $a_I(k) = \begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases}$. Thus, (721) is proven in Case 1.

Let us now consider Case 2. In this case, we have $1 < k \leq I$ and $k < I$. From $k < I$, we obtain $k \leq I - 1$ (since k and I are integers). Thus, (720) yields $a_{I-1}(k) = \begin{cases} I-1, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I-1; \\ k, & \text{if } k > I-1 \end{cases} = k - 1$ (since $1 < k \leq I - 1$) and thus $\underbrace{a_I}_{=s_{I-1} \circ a_{I-1}}(k) = (s_{I-1} \circ a_{I-1})(k) = s_{I-1}\left(\underbrace{a_{I-1}(k)}_{=k-1}\right) = s_{I-1}(k-1) = k - 1$ (since $k - 1 < k \leq I - 1$). Compared with $\begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases} = k - 1$ (since $1 < k \leq I$), this yields $a_I(k) = \begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases}$. Thus, (721) is proven in Case 2.

Let us now consider Case 3. In this case, we have $1 < k \leq I$ and $k \geq I$. Combining $k \leq I$ with $k \geq I$, we obtain $k = I > I - 1$. Thus, (720) yields $a_{I-1}(k) = \begin{cases} I-1, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I-1; \\ k, & \text{if } k > I-1 \end{cases} = k$

Next, for every $m \in \{0, 1, \ldots, n\}$, set

$$b_m = a_1 \circ a_2 \circ \cdots \circ a_m \in S_n.$$

As a consequence, $b_0 = a_1 \circ a_2 \circ \cdots \circ a_0 = (\text{a composition of 0 maps}) = \text{id}$ and $b_n = a_1 \circ a_2 \circ \cdots \circ a_n$. We claim that

$$b_m(k) = \begin{cases} m+1-k, & \text{if } k \leq m; \\ k, & \text{if } k > m \end{cases} \tag{722}$$

for every $m \in \{0, 1, \ldots, n\}$ and $k \in \{1, 2, \ldots, n\}$.

[Again, one can prove (722) either formally by induction on m, or more intuitively by tracking what happens to k under the maps a_m, a_{m-1}, ..., a_1 when these maps are applied one after the other. This time the informal way is a bit tricky, so let us show the formal one in all its glory. (You do not ever need to write a proof in this level of detail unless you are talking to a computer.)

Proof of (722): We shall prove (722) by induction on m:

Induction base: For every $k \in \{1, 2, \ldots, n\}$, we have $\underbrace{b_0}_{=\text{id}}(k) = \text{id}(k) = k = \begin{cases} 0+1-k, & \text{if } k \leq 0; \\ k, & \text{if } k > 0 \end{cases}$ (since $\begin{cases} 0+1-k, & \text{if } k \leq 0; \\ k, & \text{if } k > 0 \end{cases} = k$ (since $k > 0$)). In other words, (722) holds for $m = 0$. The induction base is thus complete.

(since $k > I-1$) and thus $\underbrace{a_I}_{=s_{I-1} \circ a_{I-1}}(k) = (s_{I-1} \circ a_{I-1})(k) = s_{I-1}\left(\underbrace{a_{I-1}(k)}_{=k=I}\right) = s_{I-1}(I) = \underbrace{I}_{=k} - 1 = k-1$. Compared with $\begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases} = k-1$ (since $1 < k \leq I$), this yields $a_I(k) = \begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases}$. Thus, (721) is proven in Case 3.

Let us finally consider Case 4. In this case, we have $k > I$. Hence, $k > I > I-1$. Thus, (720) yields $a_{I-1}(k) = \begin{cases} I-1, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I-1; \\ k, & \text{if } k > I-1 \end{cases} = k$ (since $k > I-1$) and thus $\underbrace{a_I}_{=s_{I-1} \circ a_{I-1}}(k) = (s_{I-1} \circ a_{I-1})(k) = s_{I-1}\left(\underbrace{a_{I-1}(k)}_{=k}\right) = s_{I-1}(k) = k$ (since $k > I$). Compared with $\begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases} = k$ (since $k > I$), this yields $a_I(k) = \begin{cases} I, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq I; \\ k, & \text{if } k > I \end{cases}$. Thus, (721) is proven in Case 4.

Hence, (721) is proven in each of the four Cases 1, 2, 3 and 4. Since these four Cases are the only possible cases, this shows that (721) always holds, qed.

Induction step: Let $M \in \{0, 1, \ldots, n\}$ be such that $M > 0$. Assume that (722) holds for $m = M - 1$. We need to show that (722) holds for $m = M$.

We have assumed that (722) holds for $m = M - 1$. In other words,

$$b_{M-1}(k) = \begin{cases} (M-1) + 1 - k, & \text{if } k \leq M - 1; \\ k, & \text{if } k > M - 1 \end{cases}$$

for every $k \in \{1, 2, \ldots, n\}$. Thus, for every $k \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned} b_{M-1}(k) &= \begin{cases} (M-1) + 1 - k, & \text{if } k \leq M - 1; \\ k, & \text{if } k > M - 1 \end{cases} \\ &= \begin{cases} M - k, & \text{if } k \leq M - 1; \\ k, & \text{if } k > M - 1 \end{cases} \end{aligned} \tag{723}$$

(since $(M-1) + 1 - k = M - k$).

The definition of b_{M-1} yields $b_{M-1} = a_1 \circ a_2 \circ \cdots \circ a_{M-1}$. The definition of b_M yields $b_M = a_1 \circ a_2 \circ \cdots \circ a_M = \underbrace{(a_1 \circ a_2 \circ \cdots \circ a_{M-1})}_{= b_{M-1}} \circ a_M = b_{M-1} \circ a_M$. Thus, we obtain

$$b_M(k) = \begin{cases} M + 1 - k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases} \tag{724}$$

for every $k \in \{1, 2, \ldots, n\}$ [348]. In other words, (722) holds for $m = M$. This completes the induction step. Thus, the induction proof of (722) is complete.]

Now, every $k \in \{1, 2, \ldots, n\}$ satisfies

$$\begin{aligned} b_n(k) &= \begin{cases} n + 1 - k, & \text{if } k \leq n; \\ k, & \text{if } k > n \end{cases} \qquad \text{(by (722), applied to } m = n) \\ &= n + 1 - k \qquad \text{(since } k \leq n) \\ &= w_0(k) \qquad \text{(since } w_0(k) = n + 1 - k \text{ (by the definition of } w_0)) . \end{aligned}$$

In other words, $b_n = w_0$, so that $w_0 = b_n = a_1 \circ a_2 \circ \cdots \circ a_n$. This proves (718). As we know, this solves Exercise 5.1 **(c)**. □

[348] *Proof of (724):* Let $k \in \{1, 2, \ldots, n\}$. We need to prove (724). We are in one of the following three cases:

Case 1: We have $k = 1$.

Case 2: We have $1 < k \leq M$.

Case 3: We have $k > M$.

Let us first consider Case 1. In this case, we have $k = 1$. Thus, $k = 1 \leq M$ (since $M \geq 1$ (since $M > 0$)). But (719) (applied to $i = M$) yields $a_M(k) = \begin{cases} M, & \text{if } k = 1; \\ k - 1, & \text{if } 1 < k \leq M; \\ k, & \text{if } k > M \end{cases} = M$ (since

7.41. Solution to Exercise 5.2

Let us first state Exercise 5.2 **(d)** as a separate result:

Proposition 7.71. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ be a permutation satisfying $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$. Then, $\sigma = \mathrm{id}$.

Proposition 7.71 essentially says that the only way to list the numbers $1, 2, \ldots, n$ in increasing order is $(1, 2, \ldots, n)$. If you think this is intuitively obvious, you are right. Let me nevertheless give two proofs (the second of which is completely formal):

$k = 1$), so that

$$\underbrace{b_M}_{=b_{M-1} \circ a_M}(k) = (b_{M-1} \circ a_M)(k) = b_{M-1}\left(\underbrace{a_M(k)}_{=M}\right) = b_{M-1}(M)$$
$$= \begin{cases} M - M, & \text{if } M \leq M-1; \\ M, & \text{if } M > M-1 \end{cases} \qquad \text{(by (723), applied to } M \text{ instead of } k\text{)}$$
$$= M \qquad (\text{since } M > M-1)$$
$$= M + 1 - \underbrace{1}_{=k} = M + 1 - k = \begin{cases} M+1-k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases}$$
$$\left(\text{since } \begin{cases} M+1-k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases} = M+1-k \text{ (because } k \leq M\text{)}\right).$$

Thus, (724) is proven in Case 1.

Let us now consider Case 2. In this case, we have $1 < k \leq M$. Now, (719) (applied to $i = M$) yields $a_M(k) = \begin{cases} M, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq M; \\ k, & \text{if } k > M \end{cases} = k - 1$ (since $1 < k \leq M$), so that

$$\underbrace{b_M}_{=b_{M-1} \circ a_M}(k) = (b_{M-1} \circ a_M)(k) = b_{M-1}\left(\underbrace{a_M(k)}_{=k-1}\right) = b_{M-1}(k-1)$$
$$= \begin{cases} M - (k-1), & \text{if } k-1 \leq M-1; \\ k-1, & \text{if } k-1 > M-1 \end{cases} \qquad \left(\begin{array}{c} \text{by (723), applied to } k-1 \\ \text{instead of } k \end{array}\right)$$
$$= M - (k-1) \qquad (\text{since } k-1 \leq M-1 \text{ (since } k \leq M\text{)})$$
$$= M + 1 - k = \begin{cases} M+1-k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases}$$
$$\left(\text{since } \begin{cases} M+1-k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases} = M+1-k \text{ (because } k \leq M\text{)}\right).$$

Thus, (724) is proven in Case 2.

Let us finally consider Case 3. In this case, we have $k > M$. Now, (719) (applied to $i = M$)

First proof of Proposition 7.71. First of all, every $k \in \{1, 2, \ldots, n-1\}$ satisfies $\sigma(k) < \sigma(k+1)$ [349]. In other words, $\sigma(1) < \sigma(2) < \cdots < \sigma(n)$.

Let $i \in \{1, 2, \ldots, n\}$. Then, $\sigma(i)$ is the i-th smallest among the numbers $\sigma(1), \sigma(2), \ldots, \sigma(n)$ (because $\sigma(1) < \sigma(2) < \cdots < \sigma(n)$). But since the numbers $\sigma(1), \sigma(2), \ldots, \sigma(n)$ are just the numbers $1, 2, \ldots, n$ (possibly in a different order)[350], it is clear that the i-th smallest among these numbers is i. Thus, $\sigma(i) = i$ (since $\sigma(i)$ is the i-th smallest among the numbers $\sigma(1), \sigma(2), \ldots, \sigma(n)$). Hence, $\sigma(i) = i = \operatorname{id}(i)$.

Let us now forget that we fixed i. Thus, we have shown that $\sigma(i) = \operatorname{id}(i)$ for every $i \in \{1, 2, \ldots, n\}$. In other words, $\sigma = \operatorname{id}$. Proposition 7.71 is thus proven. □

Second proof of Proposition 7.71. We shall show that

$$\sigma(i) = i \qquad \text{for every } i \in \{1, 2, \ldots, n\}. \tag{725}$$

[*Proof of (725):* We shall prove (725) by strong induction over i. Thus, we fix some $I \in \{1, 2, \ldots, n\}$, and we assume that (725) is proven for every $i < I$. We then have to prove that (725) holds for $i = I$.

We have assumed that (725) is proven for every $i < I$. In other words,

$$\sigma(i) = i \qquad \text{for every } i \in \{1, 2, \ldots, n\} \text{ satisfying } i < I. \tag{726}$$

We assume (for the sake of contradiction) that $\sigma(I) \neq I$. But σ is a permutation, and thus injective. Hence, from $\sigma(I) \neq I$, we obtain $\sigma(\sigma(I)) \neq \sigma(I)$. But if

---

yields $a_M(k) = \begin{cases} M, & \text{if } k = 1; \\ k-1, & \text{if } 1 < k \leq M; \\ k, & \text{if } k > M \end{cases} = k$ (since $k > M$), so that

$$\begin{aligned}
\underbrace{b_M}_{=b_{M-1} \circ a_M}(k) &= (b_{M-1} \circ a_M)(k) = b_{M-1}\left(\underbrace{a_M(k)}_{=k}\right) = b_{M-1}(k) \\
&= \begin{cases} M-k, & \text{if } k \leq M-1; \\ k, & \text{if } k > M-1 \end{cases} \qquad (\text{by (723)}) \\
&= k \qquad (\text{since } k > M > M-1) \\
&= \begin{cases} M+1-k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases} \\
&\qquad \left(\text{since } \begin{cases} M+1-k, & \text{if } k \leq M; \\ k, & \text{if } k > M \end{cases} = k \text{ (because } k > M)\right).
\end{aligned}$$

Thus, (724) is proven in Case 3.

Hence, (724) is proven in each of the three Cases 1, 2 and 3. Since these three Cases are the only possible cases, this shows that (724) always holds, qed.

[349] *Proof.* Let $k \in \{1, 2, \ldots, n-1\}$. Then, σ is a permutation, thus injective. Hence, $\sigma(k) \neq \sigma(k+1)$ (since $k \neq k+1$). Combined with $\sigma(k) \leq \sigma(k+1)$ (since $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$), this yields $\sigma(k) < \sigma(k+1)$, qed.

[350] since σ is a permutation of $\{1, 2, \ldots, n\}$

$\sigma(I) < I$, then $\sigma(\sigma(I)) = \sigma(I)$ (by (726), applied to $i = \sigma(I)$), which contradicts $\sigma(\sigma(I)) \neq \sigma(I)$. Hence, we cannot have $\sigma(I) < I$. Thus, we have $\sigma(I) \geq I$. Combined with $\sigma(I) \neq I$, this yields $\sigma(I) > I$.

Now, let $K = \sigma^{-1}(I)$. Then, $I = \sigma(K)$, so that $\sigma(K) = I \neq \sigma(I)$ and therefore $K \neq I$. If $K < I$, then $\sigma(K) = K$ (by (726), applied to $i = K$), which contradicts $\sigma(K) = I \neq K$. Hence, we cannot have $K < I$. We thus have $I \leq K$.

Now, recall that $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$. In other words, $\sigma(a) \leq \sigma(b)$ for every two elements a and b of $\{1, 2, \ldots, n\}$ satisfying $a \leq b$. Applying this to $a = I$ and $b = K$, we obtain $\sigma(I) \leq \sigma(K)$. This contradicts $\sigma(I) > I = \sigma(K)$. This contradiction proves that our assumption (that $\sigma(I) \neq I$) was wrong. Hence, we must have $\sigma(I) = I$. In other words, (725) holds for $i = I$. This completes our inductive proof of (725).]

Now, (725) shows that every $i \in \{1, 2, \ldots, n\}$ satisfies $\sigma(i) = i = \text{id}(i)$. In other words, $\sigma = \text{id}$. Proposition 7.71 is solved again. □

For future use, let us record an easy consequence of Proposition 7.71:

Corollary 7.72. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ be a permutation satisfying $\ell(\sigma) = 0$. Then, $\sigma = \text{id}$.

Proof of Corollary 7.72. Let $k \in \{1, 2, \ldots, n-1\}$.

Assume (for the sake of contradiction) that $\sigma(k) > \sigma(k+1)$. Then, $(k, k+1)$ is a pair of integers satisfying $1 \leq k < k+1 \leq n$ and $\sigma(k) > \sigma(k+1)$. In other words, $(k, k+1)$ is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. In other words, $(k, k+1)$ is an inversion of σ (by the definition of an "inversion"). Thus, the permutation σ has at least one inversion (namely, $(k, k+1)$).

But the number of inversions of σ is $\ell(\sigma) = 0$; in other words, σ has no inversions. This contradicts the fact that σ has at least one inversion. This contradiction proves that our assumption (that $\sigma(k) > \sigma(k+1)$) was wrong. Hence, we have $\sigma(k) \leq \sigma(k+1)$.

Now, let us forget that we fixed k. We thus have shown that $\sigma(k) \leq \sigma(k+1)$ for every $k \in \{1, 2, \ldots, n-1\}$. Thus, $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$. Therefore, Proposition 7.71 shows that $\sigma = \text{id}$. This proves Corollary 7.72. □

Now, we come to the actual solution of Exercise 5.2.

Solution to Exercise 5.2. Exercise 5.2 **(d)** follows immediately from Proposition 7.71. We shall next prove part **(f)** of the exercise, then part **(a)**, then part **(e)**, and then the remaining three parts.

Before we come to the actual solution, let us introduce one more notation.

For every $\sigma \in S_n$, let $\text{Inv}(\sigma)$ be the set of all inversions of the permutation σ.

Thus, for every $\sigma \in S_n$, we have

$$\begin{aligned} \ell(\sigma) &= (\text{the number of inversions of } \sigma) \qquad (\text{by the definition of } \ell(\sigma)) \\ &= (\text{the number of elements of } \operatorname{Inv}(\sigma)) \\ &\qquad (\text{since } \operatorname{Inv}(\sigma) \text{ is the set of all inversions of } \sigma) \\ &= |\operatorname{Inv}(\sigma)|. \end{aligned} \tag{727}$$

(d) Let $\sigma \in S_n$ be a permutation satisfying $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$. Then, Proposition 7.71 shows that $\sigma = \operatorname{id}$. This solves Exercise 5.2 **(d)**.

(f) Let $\sigma \in S_n$. For every $(i,j) \in \operatorname{Inv}(\sigma)$, we have $(\sigma(j), \sigma(i)) \in \operatorname{Inv}(\sigma^{-1})$ [351]. Hence, we can define a map

$$\Phi : \operatorname{Inv}(\sigma) \to \operatorname{Inv}\left(\sigma^{-1}\right), \qquad (i,j) \mapsto (\sigma(j), \sigma(i)).$$

This map Φ is injective[352]. Thus, we have found an injective map from $\operatorname{Inv}(\sigma)$ to $\operatorname{Inv}(\sigma^{-1})$. Conversely, $|\operatorname{Inv}(\sigma)| \leq |\operatorname{Inv}(\sigma^{-1})|$. But $\ell(\sigma^{-1}) = |\operatorname{Inv}(\sigma^{-1})|$ (by (727), applied to σ^{-1} instead of σ). Now, (727) yields $\ell(\sigma) = |\operatorname{Inv}(\sigma)| \leq |\operatorname{Inv}(\sigma^{-1})| = \ell(\sigma^{-1})$.

Now, let us forget that we fixed σ. We thus have proven that

$$\ell(\sigma) \leq \ell\left(\sigma^{-1}\right) \qquad \text{for every } \sigma \in S_n. \tag{728}$$

Now, let $\sigma \in S_n$ again. We can apply (728) to σ^{-1} instead of σ, and thus obtain $\ell(\sigma^{-1}) \leq \ell\left(\underbrace{\left(\sigma^{-1}\right)^{-1}}_{=\sigma}\right) = \ell(\sigma)$. Combined with (728), this yields $\ell(\sigma) = \ell(\sigma^{-1})$. This solves Exercise 5.2 **(f)**.

(a) As I warned above, this solution will be a tedious formalization of the argument sketched in Example 5.11.

Let us first show a very simple fact: If u and v are two integers such that $1 \leq u < v \leq n$, and if $k \in \{1, 2, \ldots, n-1\}$ is such that $(u, v) \neq (k, k+1)$, then

$$s_k(u) < s_k(v) \tag{729}$$

[351] *Proof.* Let $(i,j) \in \operatorname{Inv}(\sigma)$. Then, (i,j) is an inversion of σ (since $\operatorname{Inv}(\sigma)$ is the set of all inversions of σ). In other words, (i,j) is a pair of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$ (by the definition of an "inversion of σ"). Hence, $\sigma(j) < \sigma(i)$, so that $1 \leq \sigma(j) < \sigma(i) \leq n$; also, $\sigma^{-1}(\sigma(i)) = i < j = \sigma^{-1}(\sigma(j))$, so that $\sigma^{-1}(\sigma(j)) > \sigma^{-1}(\sigma(i))$. Therefore, $(\sigma(j), \sigma(i))$ is a pair of integers (u,v) satisfying $1 \leq u < v \leq n$ and $\sigma^{-1}(u) > \sigma^{-1}(v)$ (since $1 \leq \sigma(j) < \sigma(i) \leq n$ and $\sigma^{-1}(\sigma(j)) > \sigma^{-1}(\sigma(i))$). In other words, $(\sigma(j), \sigma(i))$ is an inversion of σ^{-1} (because inversions of σ^{-1} are defined as pairs of integers (u,v) satisfying $1 \leq u < v \leq n$ and $\sigma^{-1}(u) > \sigma^{-1}(v)$). In other words, $(\sigma(j), \sigma(i)) \in \operatorname{Inv}(\sigma^{-1})$ (since $\operatorname{Inv}(\sigma^{-1})$ is the set of all inversions of σ^{-1}), qed.

[352] *Proof.* We simply need to prove that an element $(i,j) \in \operatorname{Inv}(\sigma)$ can be reconstructed from its image $(\sigma(j), \sigma(i))$. But this is easy: If you know $(\sigma(j), \sigma(i))$, then you know $\sigma(j)$ and $\sigma(i)$, and therefore also i (since $i = \sigma^{-1}(\sigma(i))$) and j (since $j = \sigma^{-1}(\sigma(j))$), and thus also (i,j).

[353].

Recall that $s_i^2 = \operatorname{id}$ for each $i \in \{1, 2, \ldots, n-1\}$. Applying this to $i = k$, we obtain $s_k^2 = \operatorname{id}$; thus, $s_k \circ s_k = s_k^2 = \operatorname{id}$ and therefore $s_k^{-1} = s_k$.

We shall now show that

$$\begin{aligned} &\operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} \\ &= \operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\} \end{aligned} \tag{732}$$

for every $\sigma \in S_n$ and $k \in \{1, 2, \ldots, n-1\}$.

[Notice that we do not necessarily have $\left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \in \operatorname{Inv}(\sigma)$; nor do we always have $\left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \in \operatorname{Inv}(s_k \circ \sigma)$. In fact, for each given σ and k, exactly one of these two statements holds. But we can form the difference $A \setminus B$

[353] *Proof of (729):* We can prove (729) by analyzing three cases (Case 1 is when $u = k$, Case 2 is when $u = k+1$, and Case 3 is when $u \notin \{k, k+1\}$), each of which can be split into three subcases (Subcase 1 is when $v = k$, Subcase 2 is when $v = k+1$, and Subcase 3 is when $v \notin \{k, k+1\}$). These are (altogether) nine subcases, but four of them (namely, Subcases 1 and 2 in Case 1, and Subcases 1 and 2 in Case 2) are impossible (because $u < v$ and $(u, v) \neq (k, k+1)$), and the proof of (729) is easy in the remaining five subcases.

Here is a smarter way to prove (729): Let u and v be two integers such that $1 \leq u < v \leq n$. Let $k \in \{1, 2, \ldots, n-1\}$ be such that $(u, v) \neq (k, k+1)$. We need to prove (729). Indeed, assume the contrary. Thus, $s_k(u) \geq s_k(v)$.

But $u < v$ and thus $u \neq v$. The map s_k is a permutation, thus bijective, and therefore injective. Hence, $s_k(u) \neq s_k(v)$ (since $u \neq v$). Combined with $s_k(u) \geq s_k(v)$, this yields $s_k(u) > s_k(v)$. Thus, $s_k(u) \geq s_k(v) + 1$ (since $s_k(u)$ and $s_k(v)$ are integers), so that $s_k(v) + 1 \leq s_k(u)$.

We have $u < v$ and thus $u + 1 \leq v$ (since u and v are integers).

Recall that s_k is the permutation in S_n which switches k and $k+1$, while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged. Hence,

$$s_k(p) \leq p + 1 \qquad \text{for every } p \in \{1, 2, \ldots, n\}, \tag{730}$$

and this inequality becomes an equality only for $p = k$. For the same reason, we have

$$s_k(p) \geq p - 1 \qquad \text{for every } p \in \{1, 2, \ldots, n\}, \tag{731}$$

and this inequality becomes an equality only for $p = k+1$.

Applying (730) to $p = u$, we obtain $s_k(u) \leq u + 1$. Applying (731) to $p = v$, we obtain $s_k(v) \geq v - 1$, so that $v - 1 \leq s_k(v)$ and thus $v \leq s_k(v) + 1 \leq s_k(u) \leq u + 1 \leq v$.

Combining $v \leq s_k(v) + 1$ with $s_k(v) + 1 \leq v$, we obtain $v = s_k(v) + 1$, so that $s_k(v) = v - 1$. In other words, $s_k(p) = p - 1$ holds for $p = v$. But recall that the inequality (731) becomes an equality only for $p = k+1$. In other words, $s_k(p) = p - 1$ holds only for $p = k+1$. Applying this to $p = v$, we obtain $v = k+1$ (since $s_k(p) = p - 1$ holds for $p = v$).

Combining $s_k(u) \leq u + 1$ with $u + 1 \leq v \leq s_k(u)$, we obtain $s_k(u) = u + 1$. In other words, $s_k(p) = p + 1$ holds for $p = u$. But recall that the inequality (730) becomes an equality only for $p = k$. In other words, $s_k(p) = p + 1$ holds only for $p = k$. Applying this to $p = u$, we obtain $u = k$ (since $s_k(p) = p + 1$ holds for $p = u$).

Now, $\left(\underbrace{u}_{=k}, \underbrace{v}_{=k+1} \right) = (k, k+1)$ contradicts $(u, v) \neq (k, k+1)$. This contradiction proves that our assumption was wrong. Hence, (729) is proven.

of two sets A and B even if B is not a subset of A, so the statement (732) still makes sense.]

[*Proof of (732):* Let $\sigma \in S_n$ and $k \in \{1, 2, \ldots, n-1\}$. Let $(i, j) \in \operatorname{Inv}(\sigma) \setminus \{(\sigma^{-1}(k+1), \sigma^{-1}(k))\}$. Thus, $(i, j) \in \operatorname{Inv}(\sigma)$ and $(i, j) \neq (\sigma^{-1}(k+1), \sigma^{-1}(k))$. Therefore, $(\sigma(j), \sigma(i)) \neq (k, k+1)$ [354].

We have $(i, j) \in \operatorname{Inv}(\sigma)$. In other words, (i, j) is an inversion of σ. In other words, (i, j) is a pair of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. Now, $\sigma(j) < \sigma(i)$, so that $1 \leq \sigma(j) < \sigma(i) \leq n$. Also, as we know, $(\sigma(j), \sigma(i)) \neq (k, k+1)$. Hence, $s_k(\sigma(j)) < s_k(\sigma(i))$ (by (729), applied to $u = \sigma(j)$ and $v = \sigma(i)$). Thus, $(s_k \circ \sigma)(j) = s_k(\sigma(j)) < s_k(\sigma(i)) = (s_k \circ \sigma)(i)$, hence $(s_k \circ \sigma)(i) > (s_k \circ \sigma)(j)$. Hence, (i, j) is a pair of integers satisfying $1 \leq i < j \leq n$ and $(s_k \circ \sigma)(i) > (s_k \circ \sigma)(j)$. In other words, (i, j) is an inversion of $s_k \circ \sigma$ (by the definition of an inversion). In other words, $(i, j) \in \operatorname{Inv}(s_k \circ \sigma)$ (since $\operatorname{Inv}(s_k \circ \sigma)$ is defined as the set of all inversions of $s_k \circ \sigma$). Furthermore, $(i, j) \neq (\sigma^{-1}(k), \sigma^{-1}(k+1))$ [355]. Thus, $(i, j) \in \operatorname{Inv}(s_k \circ \sigma) \setminus \{(\sigma^{-1}(k), \sigma^{-1}(k+1))\}$ (since $(i, j) \in \operatorname{Inv}(s_k \circ \sigma)$ and $(i, j) \neq (\sigma^{-1}(k), \sigma^{-1}(k+1))$).

Now, let us forget that we fixed (i, j). We thus have shown that every $(i, j) \in \operatorname{Inv}(\sigma) \setminus \{(\sigma^{-1}(k+1), \sigma^{-1}(k))\}$ satisfies $(i, j) \in \operatorname{Inv}(s_k \circ \sigma) \setminus \{(\sigma^{-1}(k), \sigma^{-1}(k+1))\}$. In other words,

$$\operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\} \\ \subseteq \operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\}. \qquad (733)$$

Now, let $\tau = s_k \circ \sigma$. Then, $s_k \circ \underbrace{\tau}_{= s_k \circ \sigma} = \underbrace{s_k \circ s_k}_{= s_k^2 = \operatorname{id}} \circ \sigma = \operatorname{id} \circ \sigma = \sigma$. Moreover, $\tau^{-1}(k) = \sigma^{-1}(k+1)$ [356] and $\tau^{-1}(k+1) = \sigma^{-1}(k)$ [357].

---

[354] *Proof.* Assume the contrary. Thus, $(\sigma(j), \sigma(i)) = (k, k+1)$. Hence, $\sigma(j) = k$ and $\sigma(i) = k+1$. Hence, $\left(\underbrace{\sigma^{-1}(k+1)}_{\substack{= i \\ (\text{since } \sigma(i) = k+1)}}, \underbrace{\sigma^{-1}(k)}_{\substack{= j \\ (\text{since } \sigma(j) = k)}} \right) = (i, j) \neq (\sigma^{-1}(k+1), \sigma^{-1}(k))$, which is absurd. Hence, we have found a contradiction, so that our assumption was wrong, qed.

[355] *Proof.* Assume the contrary. Thus, $(i, j) = (\sigma^{-1}(k), \sigma^{-1}(k+1))$. Thus, $i = \sigma^{-1}(k)$ and $j = \sigma^{-1}(k+1)$. Hence, $\sigma(i) = k$ (since $i = \sigma^{-1}(k)$) and $\sigma(j) = k+1$ (since $j = \sigma^{-1}(k+1)$). Hence, $k+1 = \sigma(j) < \sigma(i) = k < k+1$, which is absurd. Thus, we have found a contradiction, so that our assumption must have been wrong, qed.

[356] *Proof.* We have $\underbrace{\tau}_{= s_k \circ \sigma}(\sigma^{-1}(k+1)) = (s_k \circ \sigma)(\sigma^{-1}(k+1)) = s_k \left(\underbrace{\sigma\left(\sigma^{-1}(k+1)\right)}_{= k+1} \right) = s_k(k+1) = k$ (by the definition of s_k). Thus, $\tau^{-1}(k) = \sigma^{-1}(k+1)$, qed.

[357] for similar reasons

But recall that we have proven (733). The same arguments, but carried out for τ instead of σ, show that

$$\operatorname{Inv}(\tau) \setminus \left\{ \left(\tau^{-1}(k+1), \tau^{-1}(k) \right) \right\} \subseteq \operatorname{Inv}(s_k \circ \tau) \setminus \left\{ \left(\tau^{-1}(k), \tau^{-1}(k+1) \right) \right\}.$$

Using the identities $s_k \circ \tau = \sigma$ and $\tau^{-1}(k) = \sigma^{-1}(k+1)$ and $\tau^{-1}(k+1) = \sigma^{-1}(k)$, we can rewrite this as follows:

$$\operatorname{Inv}(\tau) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} \subseteq \operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\}.$$

Since $\tau = s_k \circ \sigma$, this further rewrites as follows:

$$\operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} \subseteq \operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\}.$$

Combining this with (733), we obtain

$$\operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} = \operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\}.$$

This proves (732).]

Now, let $k \in \{1, 2, \ldots, n-1\}$.

We shall first show that for every $\sigma \in S_n$, we have

$$\ell(s_k \circ \sigma) = \ell(\sigma) + 1 \qquad \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1). \tag{734}$$

[*Proof of (734):* Let $\sigma \in S_n$. Assume that $\sigma^{-1}(k) < \sigma^{-1}(k+1)$. Then, $\left(\sigma^{-1}(k), \sigma^{-1}(k+1)\right)$ is a pair of integers satisfying $1 \leq \sigma^{-1}(k) < \sigma^{-1}(k+1) \leq n$ and $(s_k \circ \sigma)\left(\sigma^{-1}(k)\right) > (s_k \circ \sigma)\left(\sigma^{-1}(k+1)\right)$ [358]. In other words, $\left(\sigma^{-1}(k), \sigma^{-1}(k+1)\right)$ is an inversion of $s_k \circ \sigma$. In other words, $\left(\sigma^{-1}(k), \sigma^{-1}(k+1)\right) \in \operatorname{Inv}(s_k \circ \sigma)$. Hence,

$$\left| \operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} \right| = \left| \operatorname{Inv}(s_k \circ \sigma) \right| - 1 \tag{735}$$

On the other hand, $\left(\sigma^{-1}(k+1), \sigma^{-1}(k)\right)$ is not an inversion of σ (because if it was an inversion of σ, then we would have $1 \leq \sigma^{-1}(k+1) < \sigma^{-1}(k) \leq n$ and therefore $\sigma^{-1}(k+1) < \sigma^{-1}(k) < \sigma^{-1}(k+1)$, which would be absurd). In other words, $\left(\sigma^{-1}(k+1), \sigma^{-1}(k)\right) \notin \operatorname{Inv}(\sigma)$. Thus,

$$\operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\} = \operatorname{Inv}(\sigma),$$

---

[358]because $(s_k \circ \sigma)\left(\sigma^{-1}(k)\right) = s_k\left(\underbrace{\sigma\left(\sigma^{-1}(k)\right)}_{=k}\right) = s_k(k) = k+1$ and similarly $(s_k \circ \sigma)\left(\sigma^{-1}(k+1)\right) = k$, so that $(s_k \circ \sigma)\left(\sigma^{-1}(k)\right) = k+1 > k = (s_k \circ \sigma)\left(\sigma^{-1}(k+1)\right)$

so that

$$\operatorname{Inv}(\sigma) = \operatorname{Inv}(\sigma) \setminus \left\{ \left(\sigma^{-1}(k+1), \sigma^{-1}(k) \right) \right\}$$
$$= \operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} \qquad \text{(by (732))}. \qquad (736)$$

Now, (727) yields

$$\ell(\sigma) = |\operatorname{Inv}(\sigma)| = \left| \operatorname{Inv}(s_k \circ \sigma) \setminus \left\{ \left(\sigma^{-1}(k), \sigma^{-1}(k+1) \right) \right\} \right| \qquad \text{(by (736))}$$
$$= |\operatorname{Inv}(s_k \circ \sigma)| - 1 \qquad \text{(by (735))}. \qquad (737)$$

But (727) (applied to $s_k \circ \sigma$ instead of σ) yields $\ell(s_k \circ \sigma) = |\operatorname{Inv}(s_k \circ \sigma)|$. Hence, (737) becomes

$$\ell(\sigma) = \underbrace{|\operatorname{Inv}(s_k \circ \sigma)|}_{=\ell(s_k \circ \sigma)} - 1 = \ell(s_k \circ \sigma) - 1,$$

so that $\ell(s_k \circ \sigma) = \ell(\sigma) + 1$. This proves (734).]

Next, we will show that for every $\sigma \in S_n$, we have

$$\ell(s_k \circ \sigma) = \ell(\sigma) - 1 \qquad \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1). \qquad (738)$$

[*Proof of (738):* Let $\sigma \in S_n$. Assume that $\sigma^{-1}(k) > \sigma^{-1}(k+1)$. But $\sigma^{-1}(k) = (s_k \circ \sigma)^{-1}(k+1)$ [359] and $\sigma^{-1}(k+1) = (s_k \circ \sigma)^{-1}(k)$ [360]. Thus, $(s_k \circ \sigma)^{-1}(k) = \sigma^{-1}(k+1) < \sigma^{-1}(k) = (s_k \circ \sigma)^{-1}(k+1)$. Hence, we can apply (734) to $s_k \circ \sigma$ instead of σ. As a result, we obtain

$$\ell(s_k \circ s_k \circ \sigma) = \ell(s_k \circ \sigma) + 1.$$

Since $\underbrace{s_k \circ s_k}_{=s_k^2 = \operatorname{id}} \circ \sigma = \operatorname{id} \circ \sigma = \sigma$, this rewrites as $\ell(\sigma) = \ell(s_k \circ \sigma) + 1$, so that $\ell(s_k \circ \sigma) = \ell(\sigma) - 1$. This proves (738).]

Now, (294) follows immediately by combining (734) with (738).[361]

It remains to prove (293). Indeed, let $\sigma \in S_n$. Let us recall that $(\alpha \circ \beta)^{-1} = \beta^{-1} \circ \alpha^{-1}$ for any two permutations α and β in S_n. Applying this to $\alpha = s_k$ and $\beta = \sigma^{-1}$, we obtain $\left(s_k \circ \sigma^{-1}\right)^{-1} = \underbrace{\left(\sigma^{-1}\right)^{-1}}_{=\sigma} \circ \underbrace{s_k^{-1}}_{=s_k} = \sigma \circ s_k$. But Exercise 5.2 **(f)**

---

[359] since $(s_k \circ \sigma)\left(\sigma^{-1}(k)\right) = s_k\left(\underbrace{\sigma\left(\sigma^{-1}(k)\right)}_{=k}\right) = s_k(k) = k+1$

[360] for similar reasons

[361] The term "$\begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases}$" in (294) makes sense because every $\sigma \in S_n$ and every $k \in \{1, 2, \ldots, n-1\}$ satisfies exactly one of the conditions $\sigma(k) < \sigma(k+1)$ and $\sigma(k) > \sigma(k+1)$. (Indeed, $\sigma(k) = \sigma(k+1)$ is impossible, because every permutation $\sigma \in S_n$ is injective.)

yields $\ell(\sigma) = \ell(\sigma^{-1})$. Also, Exercise 5.2 **(f)** (applied to $s_k \circ \sigma^{-1}$ instead of σ) yields $\ell(s_k \circ \sigma^{-1}) = \ell\left(\underbrace{\left(s_k \circ \sigma^{-1}\right)^{-1}}_{=\sigma \circ s_k}\right) = \ell(\sigma \circ s_k)$. But applying (294) to σ^{-1} instead of σ, we obtain

$$\ell\left(s_k \circ \sigma^{-1}\right) = \begin{cases} \ell(\sigma^{-1}) + 1, & \text{if } (\sigma^{-1})^{-1}(k) < (\sigma^{-1})^{-1}(k+1); \\ \ell(\sigma^{-1}) - 1, & \text{if } (\sigma^{-1})^{-1}(k) > (\sigma^{-1})^{-1}(k+1) \end{cases}.$$

Since $\ell(s_k \circ \sigma^{-1}) = \ell(\sigma \circ s_k)$, $\ell(\sigma^{-1}) = \ell(\sigma)$ and $(\sigma^{-1})^{-1} = \sigma$, this equality rewrites as follows:

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases}.$$

This proves (293), and thus completes the solution of Exercise 5.2 **(a)**.

(e) We shall solve Exercise 5.2 **(e)** by induction over $\ell(\sigma)$:

Induction base: Exercise 5.2 **(e)** holds in the case when $\ell(\sigma) = 0$ [362]. This completes the induction base.

Induction step: Let L be a positive integer. Assume that Exercise 5.2 **(e)** is solved in the case when $\ell(\sigma) = L - 1$. We need to solve Exercise 5.2 **(e)** in the case when $\ell(\sigma) = L$.

So let $\sigma \in S_n$ be such that $\ell(\sigma) = L$. We need to show that σ can be written as a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$).

There exists a $k \in \{1, 2, \ldots, n-1\}$ such that $\sigma(k) > \sigma(k+1)$ [363]. Fix such a k, and denote it by j. Thus, j is an element of $\{1, 2, \ldots, n-1\}$ and satisfies $\sigma(j) > \sigma(j+1)$. From (293) (applied to $k = j$), we obtain

$$\begin{aligned} \ell(\sigma \circ s_j) &= \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(j) < \sigma(j+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(j) > \sigma(j+1) \end{cases} \\ &= \underbrace{\ell(\sigma)}_{=L} - 1 \qquad (\text{since } \sigma(j) > \sigma(j+1)) \\ &= L - 1. \end{aligned}$$

[362] *Proof.* Let $\sigma \in S_n$ be such that $\ell(\sigma) = 0$. We need to show that σ can be written as a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$).

Recall that the composition of 0 permutations in S_n is id (by definition).

We have $\ell(\sigma) = 0$, and thus $\sigma = \text{id}$ (by Corollary 7.72). Therefore, σ is a composition of 0 permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$) (because the composition of 0 permutations in S_n is id). In other words, σ is a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$) (since $\ell(\sigma) = 0$). Thus, Exercise 5.2 **(e)** is solved in the case when $\ell(\sigma) = 0$, qed.

[363] *Proof.* Assume the contrary. Then, every $k \in \{1, 2, \ldots, n-1\}$ satisfies $\sigma(k) \leq \sigma(k+1)$. In other words, $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$. Exercise 5.2 **(d)** yields $\sigma = \text{id}$. Hence, $\ell(\sigma) = \ell(\text{id}) = 0$, so that $0 = \ell(\sigma) = L$. This contradicts the fact that L is a positive integer. This contradiction shows that our assumption was wrong, qed.

Hence, we can apply Exercise 5.2 **(e)** to $\sigma \circ s_j$ instead of σ (because we assumed that Exercise 5.2 **(e)** is solved in the case when $\ell(\sigma) = L - 1$). As a result, we conclude that $\sigma \circ s_j$ can be written as a composition of $\ell(\sigma \circ s_j)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, $\sigma \circ s_j$ can be written as a composition of $L - 1$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$) (since $\ell(\sigma \circ s_j) = L - 1$). In other words, there exists an $(L-1)$-tuple $(k_1, k_2, \ldots, k_{L-1}) \in \{1, 2, \ldots, n-1\}^{L-1}$ such that $\sigma \circ s_j = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{L-1}}$. Consider this $(k_1, k_2, \ldots, k_{L-1})$.

We have $\sigma \circ \underbrace{s_j \circ s_j}_{=s_j^2=\operatorname{id}} = \sigma$ and thus

$$\sigma = \underbrace{\sigma \circ s_j}_{=s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{L-1}}} \circ s_j = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{L-1}} \circ s_j .$$

The right hand side of this equality is a composition of L permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). Thus, σ can be written as a composition of L permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, σ can be written as a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$) (since $\ell(\sigma) = L$). This solves Exercise 5.2 **(e)** in the case when $\ell(\sigma) = L$. The induction step is thus complete, and Exercise 5.2 **(e)** is solved by induction.

(b) From (293), we can easily conclude that

$$\ell(\sigma \circ s_k) \equiv \ell(\sigma) + 1 \mod 2 \tag{739}$$

for every $\sigma \in S_n$ and every $k \in \{1, 2, \ldots, n-1\}$.

Thus, using induction, it is easy to prove that

$$\ell\left(\sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}\right)\right) \equiv \ell(\sigma) + p \mod 2 \tag{740}$$

for every $\sigma \in S_n$, every $p \in \mathbb{N}$ and every $(k_1, k_2, \ldots, k_p) \in \{1, 2, \ldots, n-1\}^p$.

Now, let σ and τ be two permutations in S_n. Exercise 5.2 **(e)** (applied to τ instead of σ) yields that τ can be written as a composition of $\ell(\tau)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, there exists an $\ell(\tau)$-tuple $\left(k_1, k_2, \ldots, k_{\ell(\tau)}\right) \in \{1, 2, \ldots, n-1\}^{\ell(\tau)}$ such that $\tau = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{\ell(\tau)}}$. Consider this $\left(k_1, k_2, \ldots, k_{\ell(\tau)}\right)$. Then,

$$\ell\left(\sigma \circ \underbrace{\tau}_{=s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{\ell(\tau)}}}\right) = \ell\left(\sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{\ell(\tau)}}\right)\right) \equiv \ell(\sigma) + \ell(\tau) \mod 2$$

(by (740), applied to $p = \ell(\tau)$). This solves Exercise 5.2 **(b)**.

(c) The solution of Exercise 5.2 **(c)** is mostly parallel to our above solution to Exercise 5.2 **(b)**.

From (293), we can easily conclude that

$$\ell(\sigma \circ s_k) \leq \ell(\sigma) + 1 \tag{741}$$

for every $\sigma \in S_n$ and every $k \in \{1, 2, \ldots, n-1\}$.

Thus, using induction, it is easy to prove that

$$\ell\left(\sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}\right)\right) \leq \ell(\sigma) + p \tag{742}$$

for every $\sigma \in S_n$, every $p \in \mathbb{N}$ and every $(k_1, k_2, \ldots, k_p) \in \{1, 2, \ldots, n-1\}^p$.

Now, let σ and τ be two permutations in S_n. Exercise 5.2 **(e)** (applied to τ instead of σ) yields that τ can be written as a composition of $\ell(\tau)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, there exists an $\ell(\tau)$-tuple $\left(k_1, k_2, \ldots, k_{\ell(\tau)}\right) \in \{1, 2, \ldots, n-1\}^{\ell(\tau)}$ such that $\tau = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{\ell(\tau)}}$. Consider this $\left(k_1, k_2, \ldots, k_{\ell(\tau)}\right)$. Then,

$$\ell\left(\sigma \circ \underbrace{\tau}_{= s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{\ell(\tau)}}}\right) = \ell\left(\sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{\ell(\tau)}}\right)\right) \leq \ell(\sigma) + \ell(\tau)$$

(by (742), applied to $p = \ell(\tau)$). This solves Exercise 5.2 **(c)**.

(g) Exercise 5.2 **(e)** shows that σ can be written as a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, $\ell(\sigma)$ is an $N \in \mathbb{N}$ such that σ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In order to solve Exercise 5.2 **(g)**, it only remains to show that $\ell(\sigma)$ is the **smallest** such N. In other words, it remains to show that if $N \in \mathbb{N}$ is such that σ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$), then $N \geq \ell(\sigma)$.

So let $N \in \mathbb{N}$ be such that σ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, there exists an N-tuple $(k_1, k_2, \ldots, k_N) \in \{1, 2, \ldots, n-1\}^N$ such that $\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_N}$. Applying (742) to id and N instead of σ and p, we obtain

$$\ell\left(\mathrm{id} \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_N}\right)\right) \leq \underbrace{\ell(\mathrm{id})}_{=0} + N = N.$$

Since $\mathrm{id} \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_N}\right) = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_N} = \sigma$, this rewrites as $\ell(\sigma) \leq N$. In other words, $N \geq \ell(\sigma)$. This completes our solution to Exercise 5.2 **(g)**. □

Remark 7.73. The above solution to Exercise 5.2 owes most of its length to my attempts at being precise. As Pascal said, "I have made this longer than usual because I have not had time to make it shorter". The proofs are not, per se, difficult, but this is combinatorics, and proofs in combinatorics often have to walk a tightrope between being unreadably long and unreadably terse.

Most parts of Exercise 5.2 can be proven in more than just one way. Let me briefly mention an alternative proof for parts **(b)** and **(c)**. Namely, let us use the notation $\operatorname{Inv}(\sigma)$ for the set of all inversions of a permutation σ. Let $\sigma \in S_n$ and $\tau \in S_n$. Then, it is not hard to convince oneself that

$$\operatorname{Inv}(\sigma \circ \tau) = A \cup B,$$

where A and B are the sets defined by

$$A = \{(i,j) \mid (i,j) \in \operatorname{Inv}(\tau) \text{ and } (\tau(j), \tau(i)) \notin \operatorname{Inv}(\sigma)\};$$
$$B = \{(i,j) \mid (i,j) \notin \operatorname{Inv}(\tau) \text{ and } (\tau(i), \tau(j)) \in \operatorname{Inv}(\sigma)\}$$

(where (i,j) are subject to the condition $1 \leq i < j \leq n$ both times). (Essentially, this is because an $(i,j) \in \operatorname{Inv}(\sigma \circ \tau)$ either satisfies $\tau(i) > \tau(j)$ or satisfies $\tau(i) < \tau(j)$. In the first case, this (i,j) belongs to A; in the second case, it belongs to B.) It is furthermore clear that $|A| \leq |\operatorname{Inv}(\tau)| = \ell(\tau)$ (by (727)) and $|B| \leq |\operatorname{Inv}(\sigma)| = \ell(\sigma)$ (by (727)). Hence, (727) yields $\ell(\sigma \circ \tau) = \left| \underbrace{\operatorname{Inv}(\sigma \circ \tau)}_{= A \cup B} \right| = |A \cup B| \leq \underbrace{|A|}_{\leq \ell(\tau)} + \underbrace{|B|}_{\leq \ell(\sigma)} \leq \ell(\tau) + \ell(\sigma)$. This solves Exercise 5.2 **(c)**. To solve part **(b)**, we need to take a few more steps. First, it is clear that $A \cap B = \varnothing$, so that $|A \cup B| = |A| + |B|$. Second, let us set

$$C = \left\{ (i,j) \in \{1,2,\ldots,n\}^2 \mid i < j, \, \tau(i) > \tau(j) \text{ and } \sigma(\tau(i)) < \sigma(\tau(j)) \right\}.$$

Then, it is easy to see that $C \subseteq \operatorname{Inv}(\tau)$ and $A = \operatorname{Inv}(\tau) \setminus C$, so that $|A| = |\operatorname{Inv}(\tau)| - |C|$. Moreover, if $\mathbf{t}$ denotes the permutation of $\{1,2,\ldots,n\}^2$ which sends every $(i,j) \in \{1,2,\ldots,n\}^2$ to $(\tau(i), \tau(j))$, and if $\mathbf{f}$ denotes the permutation of $\{1,2,\ldots,n\}^2$ which sends every (i,j) to (j,i), then $\mathbf{f}(C) \subseteq \mathbf{t}^{-1}(\operatorname{Inv}(\sigma))$ and $B = \mathbf{t}^{-1}(\operatorname{Inv}(\sigma)) \setminus \mathbf{f}(C)$, so that $|B| = \underbrace{\left| \mathbf{t}^{-1}(\operatorname{Inv}(\sigma)) \right|}_{= |\operatorname{Inv}(\sigma)|} - \underbrace{|\mathbf{f}(C)|}_{= |C|} = |\operatorname{Inv}(\sigma)| - |C|$.

Thus,

$$\begin{aligned}
\ell(\sigma \circ \tau) &= \left| \underbrace{\operatorname{Inv}(\sigma \circ \tau)}_{= A \cup B} \right| = |A \cup B| = \underbrace{|A|}_{= |\operatorname{Inv}(\tau)| - |C|} + \underbrace{|B|}_{= |\operatorname{Inv}(\sigma)| - |C|} \\
&= \left(\underbrace{|\operatorname{Inv}(\tau)|}_{= \ell(\tau)} - |C| \right) + \left(\underbrace{|\operatorname{Inv}(\sigma)|}_{= \ell(\sigma)} - |C| \right) = (\ell(\tau) - |C|) + (\ell(\sigma) - |C|) \\
&= \ell(\sigma) + \ell(\tau) - 2|C| \equiv \ell(\sigma) + \ell(\tau) \mod 2.
\end{aligned}$$

This solves Exercise 5.2 **(b)**.

7.42. Solution to Exercise 5.3

Solution to Exercise 5.3. We need to show that if $p \in \mathbb{N}$ and $(k_1, k_2, \ldots, k_p) \in \{1, 2, \ldots, n-1\}^p$ are such that $\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}$, then $p \equiv \ell(\sigma) \bmod 2$.

Let $p \in \mathbb{N}$ and $(k_1, k_2, \ldots, k_p) \in \{1, 2, \ldots, n-1\}^p$ be such that $\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}$. We must prove that $p \equiv \ell(\sigma) \bmod 2$.

Applying (740) to id instead of σ, we see that

$$\ell\left(\operatorname{id} \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}\right)\right) \equiv \underbrace{\ell(\operatorname{id})}_{=0} + p = p \bmod 2.$$

Since $\operatorname{id} \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}\right) = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p} = \sigma$, this rewrites as $\ell(\sigma) \equiv p \bmod 2$. In other words, $p \equiv \ell(\sigma) \bmod 2$. This completes the solution of Exercise 5.3. □

7.43. Solution to Exercise 5.4

Solution to Exercise 5.4. We have $n \geq 2$. Thus, the permutation $s_1 \in S_n$ is well-defined. (This is the permutation which switches 1 with 2 while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged.)

Let A_n denote the set of all even permutations in S_n. Let C_n denote the set of all odd permutations in S_n. The sign of a permutation in S_n is either 1 or -1 (because it is defined to be an integer power of -1), but not both. Hence, every permutation in S_n is either even or odd, but not both. In other words, we have $S_n = A_n \cup C_n$ and $A_n \cap C_n = \varnothing$. Therefore, $|S_n| = |A_n| + |C_n|$.

Now, for every $\sigma \in A_n$, we have $\sigma \circ s_k \in C_n$ [364]. Hence, we can define a map $\Phi : A_n \to C_n$ by

$$\Phi(\sigma) = \sigma \circ s_k \qquad \text{for every } \sigma \in A_n.$$

Similarly, we can define a map $\Psi : C_n \to A_n$ by

$$\Psi(\sigma) = \sigma \circ s_k \qquad \text{for every } \sigma \in C_n.$$

These two maps Φ and Ψ are mutually inverse (since every $\sigma \in S_n$ satisfies $\sigma \circ \underbrace{s_k \circ s_k}_{=s_k^2 = \operatorname{id}} = \sigma$). Therefore, the map Φ is a bijection. Thus, there exists a bijection form

[364] *Proof.* Let $\sigma \in A_n$. Thus, σ is an even permutation in S_n (since A_n is the set of all even permutations in S_n). Since σ is even, we have $(-1)^\sigma = 1$, so that $1 = (-1)^\sigma = (-1)^{\ell(\sigma)}$. Therefore, $\ell(\sigma) \equiv 0 \bmod 2$. Now, (739) yields $\ell(\sigma \circ s_k) \equiv \underbrace{\ell(\sigma)}_{\equiv 0 \bmod 2} + 1 \equiv 1 \bmod 2$, so that $(-1)^{\ell(\sigma \circ s_k)} = -1$. But now, $(-1)^{\sigma \circ s_k} = (-1)^{\ell(\sigma \circ s_k)} = -1$, so that the permutation $\sigma \circ s_k$ is odd. In other words, $\sigma \circ s_k \in C_n$ (since C_n is the set of all odd permutations in S_n), qed.

A_n to C_n (namely, Φ), so that we obtain $|A_n| = |C_n|$. Hence, $|S_n| = \underbrace{|A_n|}_{=|C_n|} + |C_n| = |C_n| + |C_n| = 2|C_n|$ and therefore $|C_n| = \dfrac{1}{2}\underbrace{|S_n|}_{=n!} = \dfrac{1}{2}n! = n!/2$. In other words, the number of odd permutations in S_n is $n!/2$. Similarly, the number of even permutations in S_n is $n!/2$. Exercise 5.4 is solved. □

7.44. Solution to Exercise 5.5

Solution to Exercise 5.5. The solution of Exercise 5.5 **(a)** is completely analogous to the solution of Exercise 5.1 **(a)**; it can be obtained from the latter by replacing S_n by S_∞, replacing $\{1, 2, \ldots, n\}$ by $\{1, 2, 3, \ldots\}$, and replacing $\{1, 2, \ldots, n-2\}$ by $\{1, 2, 3, \ldots\}$.

As for Exercise 5.5 **(b)**, we again omit the solution, because it follows from an exercise that will be solved below (Exercise 5.6 **(e)**). □

7.45. Solution to Exercise 5.6

Solution to Exercise 5.6. A solution of Exercise 5.6 can be obtained by copying our above solution of Exercise 5.2 almost verbatim, occasionally doing some replacements (e.g., we have to replace S_n by $S_{(\infty)}$, to replace $1 \leq i < j \leq n$ by $1 \leq i < j$, to replace $\{1, 2, \ldots, n-1\}$ by $\{1, 2, 3, \ldots\}$, and to replace $\{1, 2, \ldots, n\}$ by $\{1, 2, 3, \ldots\}$). We leave the straightforward changes to the reader. □

7.46. Solution to Exercise 5.7

Solution to Exercise 5.7. A solution of Exercise 5.7 can be obtained by copying our above solution of Exercise 5.3 almost verbatim, occasionally doing some replacements (e.g., we have to replace S_n by $S_{(\infty)}$, to replace $\{1, 2, \ldots, n-1\}$ by $\{1, 2, 3, \ldots\}$, and to replace $\{1, 2, \ldots, n\}$ by $\{1, 2, 3, \ldots\}$). We leave the straightforward changes to the reader. □

7.47. Solution to Exercise 5.8

Solution to Exercise 5.8. In the following, "path" will always mean "path on the (undirected) n-th right Bruhat graph". Hence, we need to prove that $\ell\left(\sigma^{-1} \circ \tau\right)$ is the smallest length of a path between σ and τ.

We write any path as the tuple consisting of its vertices (from its beginning to its end).[365]

Let $L = \ell\left(\sigma^{-1} \circ \tau\right)$. We shall first show that there exists a path of length L between σ and τ.

[365] This is legitimate, because the (undirected) n-th right Bruhat graph does not have multiple edges.

Indeed, Exercise 5.2 **(g)** (applied to $\sigma^{-1} \circ \tau$ instead of σ) yields that $\ell\left(\sigma^{-1} \circ \tau\right)$ is the smallest $N \in \mathbb{N}$ such that $\sigma^{-1} \circ \tau$ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). Since $L = \ell\left(\sigma^{-1} \circ \tau\right)$, we can rewrite this as follows: L is the smallest $N \in \mathbb{N}$ such that $\sigma^{-1} \circ \tau$ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). Hence, $\sigma^{-1} \circ \tau$ can be written as a composition of L permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, there exists an L-tuple $(j_1, j_2, \ldots, j_L) \in \{1, 2, \ldots, n-1\}^L$ such that $\sigma^{-1} \circ \tau = s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_L}$. Consider this $(j_1, j_2, \ldots, j_L)$.

We have $\underbrace{\sigma \circ \sigma^{-1}}_{=\text{id}} \circ \tau = \tau$, so that $\tau = \sigma \circ \underbrace{\sigma^{-1} \circ \tau}_{= s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_L}} = \sigma \circ \left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_L}\right)$. Now, for every $p \in \{0, 1, \ldots, L\}$, define a permutation $\gamma_p \in S_n$ by

$$\gamma_p = \sigma \circ \left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_p}\right).$$

Thus, $\gamma_0 = \sigma \circ \underbrace{\left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_0}\right)}_{=(\text{a composition of 0 permutations})=\text{id}} = \sigma$ and $\gamma_L = \sigma \circ \left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_L}\right) = \tau$.

Now, for every $i \in \{1, 2, \ldots, L\}$, the vertices γ_i and γ_{i-1} of the (undirected) n-th right Bruhat graph are adjacent[366]. Hence, the vertices $\gamma_0, \gamma_1, \ldots, \gamma_L$ form a path. This path connects σ to τ (since it begins at $\gamma_0 = \sigma$ and ends at $\gamma_L = \tau$), and has length L. Thus, there exists a path between σ and τ which has length L (namely, the path formed by the vertices $\gamma_0, \gamma_1, \ldots, \gamma_L$).

We shall now show that L is the smallest length of a path between σ and τ. Indeed, let $\mathbf{d}$ be any path between σ and τ. We shall show that the length of $\mathbf{d}$ is $\geq L$.

The path $\mathbf{d}$ is a path between σ and τ. Hence, we can write the path $\mathbf{d}$ in the form $\mathbf{d} = (\delta_0, \delta_1, \ldots, \delta_M)$ for some $\delta_0, \delta_1, \ldots, \delta_M \in S_n$ with $\delta_0 = \sigma$ and $\delta_M = \tau$. Consider these $\delta_0, \delta_1, \ldots, \delta_M \in S_n$.

For every $i \in \{1, 2, \ldots, M\}$, there exists a $k \in \{1, 2, \ldots, n-1\}$ such that $\delta_i = \delta_{i-1} \circ s_k$ [367]. We denote this k by k_i. Thus, for every $i \in \{1, 2, \ldots, M\}$, we have defined a $k_i \in \{1, 2, \ldots, n-1\}$ such that $\delta_i = \delta_{i-1} \circ s_{k_i}$.

[366]*Proof.* Let $i \in \{1, 2, \ldots, L\}$. Then, the definition of γ_{i-1} yields $\gamma_{i-1} = \sigma \circ \left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_{i-1}}\right)$, whereas the definition of γ_i yields

$$\gamma_i = \sigma \circ \underbrace{\left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_i}\right)}_{=\left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_{i-1}}\right) \circ s_{j_i}} = \underbrace{\sigma \circ \left(s_{j_1} \circ s_{j_2} \circ \cdots \circ s_{j_{i-1}}\right)}_{=\gamma_{i-1}} \circ s_{j_i} = \gamma_{i-1} \circ s_{j_i}.$$

Therefore, there exists a $k \in \{1, 2, \ldots, n-1\}$ such that $\gamma_i = \gamma_{i-1} \circ s_k$ (namely, $k = j_i$). In other words, the vertices γ_i and γ_{i-1} of the (undirected) n-th right Bruhat graph are adjacent (by the definition of the edges of this graph). Qed.

[367]*Proof.* Let $i \in \{1, 2, \ldots, M\}$. Then, the vertices δ_i and δ_{i-1} of the (undirected) n-th right Bruhat graph are adjacent (because they are two consecutive vertices on the path $(\delta_0, \delta_1, \ldots, \delta_M) = \mathbf{d}$). In other words, there exists a $k \in \{1, 2, \ldots, n-1\}$ such that $\delta_i = \delta_{i-1} \circ s_k$ (by the definition of the edges of this graph). Qed.

Now, every $j \in \{0, 1, \ldots, M\}$ satisfies

$$\delta_j = \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_j} \right) \tag{743}$$

[368]. Applying this to $j = M$, we obtain

$$\delta_M = \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_M} \right).$$

Compared with $\delta_M = \tau$, this yields

$$\tau = \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_M} \right),$$

so that

$$\sigma^{-1} \circ \underbrace{\tau}_{= \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_M} \right)} = \underbrace{\sigma^{-1} \circ \sigma}_{= \mathrm{id}} \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_M} \right) = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_M}.$$

Therefore, the permutation $\sigma^{-1} \circ \tau$ can be written as a composition of M permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$) (namely, of the M permutations s_{k_1}, s_{k_2}, ..., s_{k_M}).

Now, we recall that L is the smallest $N \in \mathbb{N}$ such that $\sigma^{-1} \circ \tau$ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). Hence, if $N \in \mathbb{N}$ is such that $\sigma^{-1} \circ \tau$ can be written as a composition of N permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$), then $N \geq L$. We can apply this to $N = M$ (because $\sigma^{-1} \circ \tau$ can be written as a composition of M permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$)), and thus obtain $M \geq L$.

But the length of the path $\mathbf{d}$ is M (since $\mathbf{d} = (\delta_0, \delta_1, \ldots, \delta_M)$). Hence, the length of the path $\mathbf{d}$ is $\geq L$ (since $M \geq L$).

---

[368] *Proof of (743):* We shall prove (743) by induction over j:

Induction base: We have $\delta_0 = \sigma$. Compared with $\sigma \circ \underbrace{\left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_0} \right)}_{= (\text{a composition of } 0 \text{ permutations}) = \mathrm{id}} = \sigma$, this yields $\delta_0 = \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_0} \right)$. In other words, (743) holds for $j = 0$. This completes the induction base.

Induction step: Let $J \in \{0, 1, \ldots, M\}$ be positive. Assume that (743) holds for $j = J - 1$. We need to show that (743) holds for $j = J$.

We have $\delta_J = \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{J-1}} \right)$ (since (743) holds for $j = J - 1$). Now, recall that $\delta_i = \delta_{i-1} \circ s_{k_i}$ for every $i \in \{1, 2, \ldots, M\}$. Applying this to $i = J$, we obtain

$$\begin{aligned} \delta_J &= \underbrace{\delta_{J-1}}_{= \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{J-1}} \right)} \circ s_{k_J} \qquad (\text{since } J \in \{1, 2, \ldots, M\} \ (\text{since } J \in \{0, 1, \ldots, M\} \text{ is positive})) \\ &= \sigma \circ \underbrace{\left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_{J-1}} \right) \circ s_{k_J}}_{= s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_J}} = \sigma \circ \left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_J} \right). \end{aligned}$$

In other words, (743) holds for $j = J$. This completes the induction step. Thus, (743) is proven by induction.

Let us now forget that we fixed $\mathbf{d}$. We thus have shown that if $\mathbf{d}$ is any path between σ and τ, then the length of the path $\mathbf{d}$ is $\geq L$. In other words, every path between σ and τ has length $\geq L$.

Altogether, we have proven the following two statements:

- There exists a path of length L between σ and τ.
- Every path between σ and τ has length $\geq L$.

Therefore, L is the smallest length of a path between σ and τ. In other words, $\ell\left(\sigma^{-1} \circ \tau\right)$ is the smallest length of a path between σ and τ (since $L = \ell\left(\sigma^{-1} \circ \tau\right)$). Exercise 5.8 is solved. $\square$

7.48. Solution to Exercise 5.9

7.48.1. Preparations

Before we solve Exercise 5.9, let us prepare with some simple lemmas and notations. The following notation will be used throughout Section 7.48:

Definition 7.74. Whenever m is an integer, we shall use the notation $[m]$ for the set $\{1, 2, \ldots, m\}$. (This is an empty set when $m \leq 0$.)

Notice that if a and b are two integers satisfying $a \leq b$, then $[a] \subseteq [b]$.

The following is a simple property of the permutations $t_{i,j}$ defined in Definition 5.30:

Lemma 7.75. Let $n \in \mathbb{N}$. Let i and j be two elements of $[n]$.

(a) We have $t_{i,j}(i) = j$.

(b) We have $t_{i,j}(j) = i$.

(c) We have $t_{i,j}(k) = k$ for each $k \in [n] \setminus \{i, j\}$.

(d) We have $t_{i,j} \circ t_{i,j} = \operatorname{id}$.

Proof of Lemma 7.75. Lemma 7.75 follows from the definition of $t_{i,j}$ (given in Definition 5.29 and in Definition 5.30). $\square$

Lemma 7.76. Let $n \in \mathbb{N}$. Let $k \in [n]$. Let $\sigma \in S_n$ be such that

$$(\sigma(i) = i \text{ for each } i \in \{k+1, k+2, \ldots, n\}). \qquad (744)$$

Let $g = \sigma^{-1}(k)$. Then:

(a) We have $g \in [k]$.

(b) We have $\left(\sigma \circ t_{k,g}\right)(i) = i$ for each $i \in \{k, k+1, \ldots, n\}$.

Proof of Lemma 7.76. We have $\sigma \in S_n$. In other words, σ is a permutation of $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\}$), that is, a bijection $[n] \to [n]$. Hence, the inverse σ^{-1} of σ is well-defined.

From $g = \sigma^{-1}(k)$, we obtain $\sigma(g) = k$.

(a) Assume the contrary. Thus, $g \notin [k]$. Combining $g \in [n]$ with $g \notin [k]$, we obtain $g \in [n] \setminus [k] = \{k+1, k+2, \ldots, n\}$. Hence, (744) (applied to $i = g$) yields $\sigma(g) = g$. Hence, $g = \sigma(g) = k \in [k]$. This contradicts $\sigma(k) \notin [k]$. This contradiction shows that our assumption was wrong. Hence, Lemma 7.76 **(a)** is proven.

(b) Let $i \in \{k, k+1, \ldots, n\}$. We must prove that $(\sigma \circ t_{k,g})(i) = i$.

We are in one of the following two cases:

Case 1: We have $i \neq k$.

Case 2: We have $i = k$.

Let us first consider Case 1. In this case, we have $i \neq k$. Combining $i \in \{k, k+1, \ldots, n\}$ with $i \neq k$, we obtain $i \in \{k, k+1, \ldots, n\} \setminus \{k\} = \{k+1, k+2, \ldots, n\}$. Hence, (744) shows that $\sigma(i) = i$.

But $i \notin \{k, g\}$ [369]. Combining $i \in \{k, k+1, \ldots, n\} \subseteq [n]$ with $i \notin \{k, g\}$, we obtain $i \in [n] \setminus \{k, g\}$. Hence, Lemma 7.75 **(c)** (applied to k, g and i instead of i, j and k) shows that $t_{k,g}(i) = i$.

Now, $(\sigma \circ t_{k,g})(i) = \sigma\left(\underbrace{t_{k,g}(i)}_{=i}\right) = \sigma(i) = i$. Thus, $(\sigma \circ t_{k,g})(i) = i$ is proven in Case 1.

Let us now consider Case 2. In this case, we have $i = k$. Lemma 7.75 **(a)** (applied to k and g instead of i and j) yields $t_{k,g}(k) = g$. Thus, $(\sigma \circ t_{k,g})\left(\underbrace{i}_{=k}\right) = (\sigma \circ t_{k,g})(k) = \sigma\left(\underbrace{t_{k,g}(k)}_{=g}\right) = \sigma(g) = k = i$. Thus, $(\sigma \circ t_{k,g})(i) = i$ is proven in Case 2.

We now have proven $(\sigma \circ t_{k,g})(i) = i$ in each of the two Cases 1 and 2. Hence, $(\sigma \circ t_{k,g})(i) = i$ always holds. Thus, Lemma 7.76 **(b)** is proven. □

Lemma 7.77. Let $n \in \mathbb{N}$. Let $k \in \{0, 1, \ldots, n\}$ and $\sigma \in S_n$. Assume that

$$(\sigma(i) = i \text{ for each } i \in \{k+1, k+2, \ldots, n\}). \tag{745}$$

Then, there exists a k-tuple $(i_1, i_2, \ldots, i_k) \in [1] \times [2] \times \cdots \times [k]$ such that

$$\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k}.$$

[369] *Proof.* Assume the contrary. Thus, $i \in \{k, g\}$. Combining this with $i \neq k$, we obtain $i \in \{k, g\} \setminus \{k\} \subseteq \{g\}$. Thus, $i = g \in [k]$ (by Lemma 7.76 **(a)**). Hence, $i \leq k$. But $i \in \{k+1, k+2, \ldots, n\}$, so that $i \geq k+1 > k$. This contradicts $i \leq k$. This contradiction shows that our assumption was wrong. Qed.

Proof of Lemma 7.77. We shall prove Lemma 7.77 by induction over k:

Induction base: Lemma 7.77 is true when $k = 0$ [370].

Induction step: Let $K \in \{0, 1, \ldots, n\}$ be positive. Assume that Lemma 7.77 holds when $k = K - 1$. We must show that Lemma 7.77 holds when $k = K$.

We have assumed that Lemma 7.77 holds when $k = K - 1$. In other words, the following statement holds:

> *Statement 1:* Let $n \in \mathbb{N}$. Assume that $K - 1 \in \{0, 1, \ldots, n\}$. Let $\sigma \in S_n$. Assume that
>
> $$(\sigma(i) = i \text{ for each } i \in \{(K-1)+1, (K-1)+2, \ldots, n\}).$$
>
> Then, there exists a $(K-1)$-tuple $(i_1, i_2, \ldots, i_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$ such that
>
> $$\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{K-1,i_{K-1}}.$$

We must show that Lemma 7.77 holds when $k = K$. In other words, we must prove the following statement:

> *Statement 2:* Let $n \in \mathbb{N}$. Assume that $K \in \{0, 1, \ldots, n\}$. Let $\sigma \in S_n$. Assume that
>
> $$(\sigma(i) = i \text{ for each } i \in \{K+1, K+2, \ldots, n\}).$$
>
> Then, there exists a K-tuple $(i_1, i_2, \ldots, i_K) \in [1] \times [2] \times \cdots \times [K]$ such that
>
> $$\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{K,i_K}.$$

[*Proof of Statement 2:* We have $K \neq 0$ (since K is positive). Combined with $K \in \{0, 1, \ldots, n\}$, this yields $K \in \{0, 1, \ldots, n\} \setminus \{0\} = \{1, 2, \ldots, n\}$, so that $K - 1 \in \{0, 1, \ldots, n-1\} \subseteq \{0, 1, \ldots, n\}$.

We have $\sigma \in S_n$. In other words, σ is a permutation of $[n]$ (since S_n is the set of all permutations of $[n]$). Hence, σ has an inverse σ^{-1}.

Moreover, $K \in \{1, 2, \ldots, n\} = [n]$. We can thus define $g \in [n]$ by $g = \sigma^{-1}(K)$. Consider this g.

[370] *Proof.* Let n, k and σ be as in Lemma 7.77. Assume that $k = 0$. We have to prove that Lemma 7.77 holds under this assumption.

In view of $k = 0$, the assumption (745) rewrites as follows: We have $\sigma(i) = i$ for each $i \in \{1, 2, \ldots, n\}$. In other words, $\sigma = \text{id}$.

If $(i_1, i_2, \ldots, i_0) \in [1] \times [2] \times \cdots \times [0]$ is the empty 0-tuple $()$, then

$$t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{0,i_0} = (\text{empty composition of permutations}) = \text{id} = \sigma.$$

Hence, there exists a 0-tuple $(i_1, i_2, \ldots, i_0) \in [1] \times [2] \times \cdots \times [0]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{0,i_0}$ (namely, the empty 0-tuple $()$). Since $k = 0$, this rewrites as follows: There exists a k-tuple $(i_1, i_2, \ldots, i_k) \in [1] \times [2] \times \cdots \times [k]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k}$. In other words, Lemma 7.77 holds. Thus, Lemma 7.77 is true when $k = 0$.

Lemma 7.76 **(a)** (applied to $k = K$) yields $g \in [K]$.

Define a permutation $\tau \in S_n$ by $\tau = \sigma \circ t_{K,g}$. Lemma 7.76 **(b)** (applied to $k = K$) yields that $(\sigma \circ t_{K,g})(i) = i$ for each $i \in \{K, K+1, \ldots, n\}$. In other words, $\tau(i) = i$ for each $i \in \{K, K+1, \ldots, n\}$ (since $\tau = \sigma \circ t_{K,g}$). In other words, $\tau(i) = i$ for each $i \in \{(K-1)+1, (K-1)+2, \ldots, n\}$ (since $K = (K-1)+1$). Hence, Statement 1 (applied to τ instead of σ) shows that there exists a $(K-1)$-tuple $(i_1, i_2, \ldots, i_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$ such that

$$\tau = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{K-1,i_{K-1}}.$$

Consider this $(K-1)$-tuple $(i_1, i_2, \ldots, i_{K-1})$, and denote it by $(j_1, j_2, \ldots, j_{K-1})$. Thus, $(j_1, j_2, \ldots, j_{K-1})$ is a $(K-1)$-tuple in $[1] \times [2] \times \cdots \times [K-1]$ such that

$$\tau = t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K-1,j_{K-1}}.$$

Now,

$$\underbrace{\tau}_{=\sigma \circ t_{K,g}} \circ t_{K,g} = \sigma \circ \underbrace{t_{K,g} \circ t_{K,g}}_{\substack{=\text{id} \\ \text{(by Lemma 7.75 (d),} \\ \text{applied to } i=K \text{ and } j=g)}} = \sigma.$$

Hence,

$$\sigma = \underbrace{\tau}_{=t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K-1,j_{K-1}}} \circ t_{K,g} = \left(t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K-1,j_{K-1}}\right) \circ t_{K,g}. \qquad (746)$$

Now, let us extend our $(K-1)$-tuple $(j_1, j_2, \ldots, j_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$ to a K-tuple $(j_1, j_2, \ldots, j_K) \in [1] \times [2] \times \cdots \times [K]$ by setting $j_K = g$. (This is well-defined, since $g \in [K]$.) Then,

$$t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K,j_K} = \left(t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K-1,j_{K-1}}\right) \circ \underbrace{t_{K,j_K}}_{\substack{=t_{K,g} \\ \text{(since } j_K = g)}}$$
$$= \left(t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K-1,j_{K-1}}\right) \circ t_{K,g}.$$

Comparing this with (746), we find $\sigma = t_{1,j_1} \circ t_{2,j_2} \circ \cdots \circ t_{K,j_K}$. Thus, there exists a K-tuple $(i_1, i_2, \ldots, i_K) \in [1] \times [2] \times \cdots \times [K]$ such that

$$\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{K,i_K}$$

(namely, $(i_1, i_2, \ldots, i_K) = (j_1, j_2, \ldots, j_K)$). This proves Statement 2.]

So Statement 2 is proven. In other words, Lemma 7.77 holds when $k = K$ (since Statement 2 is precisely the claim of Lemma 7.77 for $k = K$). This completes the induction step. Thus, Lemma 7.77 is proven. □

Lemma 7.78. Let $n \in \mathbb{N}$. Let $k \in \{0, 1, \ldots, n\}$ and $\sigma \in S_n$. Let $(i_1, i_2, \ldots, i_k) \in [1] \times [2] \times \cdots \times [k]$ be such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k}$. Then:

(a) We have $\sigma(i) = i$ for each $i \in \{k+1, k+2, \ldots, n\}$.

(b) If $k > 0$, then $\sigma(i_k) = k$.

Proof of Lemma 7.78. We have $(i_1, i_2, \ldots, i_k) \in [1] \times [2] \times \cdots \times [k]$. Thus,

$$i_j \in [j] \qquad \text{for each } j \in \{1, 2, \ldots, k\}. \tag{747}$$

(a) Let $i \in \{k+1, k+2, \ldots, n\}$. Thus, $i \geq k+1 > k$.

We claim that

$$\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{p,i_p}\right)(i) = i \qquad \text{for each } p \in \{0, 1, \ldots, k\}. \tag{748}$$

[*Proof of (748):* We shall prove (748) by induction over p:

Induction base: We have $\underbrace{\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{0,i_0}\right)}_{=(\text{empty composition})=\text{id}}(i) = \text{id}(i) = i$. In other words, (748) holds for $p = 0$. This completes the induction base.

Induction step: Let $q \in \{0, 1, \ldots, k\}$ be positive. Assume that (748) holds for $p = q - 1$. We must prove that (748) holds for $p = q$.

We have assumed that (748) holds for $p = q - 1$. In other words, we have $\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{q-1,i_{q-1}}\right)(i) = i$.

We have $q \in \{0, 1, \ldots, k\}$, so that $q \leq k$. Hence, $k \geq q$, so that $i > k \geq q$. Thus, $i \neq q$. Furthermore, $q \neq 0$ (since q is positive). Combining this with $q \in \{0, 1, \ldots, k\}$, we obtain $q \in \{0, 1, \ldots, k\} \setminus \{0\} = \{1, 2, \ldots, k\}$. Hence, (747) (applied to $j = q$) yields $i_q \in [q]$. Hence, $i_q \leq q$. Thus, $q \geq i_q$, so that $i > q \geq i_q$. Thus, $i \neq i_q$. Finally, $i_q \in [q] \subseteq [n]$ (since $q \leq k \leq n$). Furthermore, $q \leq k \leq n$, so that $q \in [n]$ (since q is positive).

Also, $i \in \{k+1, k+2, \ldots, n\} \subseteq \{1, 2, \ldots, n\} = [n]$. Thus, i is an element of $[n]$ other than q and i_q (since $i \neq q$ and $i \neq i_q$). In other words, $i \in [n] \setminus \{q, i_q\}$. Hence, Lemma 7.75 **(c)** (applied to q, i_q and i instead of i, j and k) shows that $t_{q,i_q}(i) = i$. Now,

$$\begin{aligned}
&\underbrace{\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{q,i_q}\right)}_{=\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{q-1,i_{q-1}}\right) \circ t_{q,i_q}}(i) \\
&= \left(\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{q-1,i_{q-1}}\right) \circ t_{q,i_q}\right)(i) = \left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{q-1,i_{q-1}}\right)\left(\underbrace{t_{q,i_q}(i)}_{=i}\right) \\
&= \left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{q-1,i_{q-1}}\right)(i) = i.
\end{aligned}$$

In other words, (748) holds for $p = q$. This completes the induction step. Thus, (748) is proven.]

Now, we can apply (748) to $p = k$ (since $k \in \{0, 1, \ldots, k\}$). We thus obtain $\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k}\right)(i) = i$. Hence,

$$\underbrace{\sigma}_{= t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k}} (i) = \left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k}\right)(i) = i.$$

This proves Lemma 7.78 **(a)**.

(b) Assume that $k > 0$. Hence, $k - 1 \in \mathbb{N}$. Hence, $k - 1 \in \{0, 1, \ldots, n\}$ (since $k - 1 \leq k \leq n$). Furthermore, $k \in \{(k-1)+1, (k-1)+2, \ldots, n\}$ (because $k = (k-1)+1 \geq (k-1)+1$ and $k \leq n$).

Define $\tau \in S_n$ by $\tau = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k-1,i_{k-1}}$. From (747), we obtain $(i_1, i_2, \ldots, i_{k-1}) \in [1] \times [2] \times \cdots \times [k-1]$. Hence, Lemma 7.78 **(a)** (applied to $k-1$, τ and k instead of k, σ and i) yields $\tau(k) = k$.

Combining $k > 0$ with $k \in \{0, 1, \ldots, n\}$, we obtain $k \in [n]$. Also, $k \in \{1, 2, \ldots, k\}$ (since $k > 0$). Hence, (747) (applied to $j = k$) yields $i_k \in [k] \subseteq [n]$ (since $k \leq n$). Thus, Lemma 7.75 **(b)** (applied to k and i_k instead of i and j) shows that $t_{k,i_k}(i_k) = k$. Now,

$$\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k} = \underbrace{\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k-1,i_{k-1}}\right)}_{=\tau} \circ t_{k,i_k} = \tau \circ t_{k,i_k}.$$

Applying both sides of this equality to i_k, we find

$$\sigma(i_k) = \left(\tau \circ t_{k,i_k}\right)(i_k) = \tau\left(\underbrace{t_{k,i_k}(i_k)}_{=k}\right) = \tau(k) = k.$$

This proves Lemma 7.78 **(b)**. □

Lemma 7.79. Let $n \in \mathbb{N}$. Let $k \in \{0, 1, \ldots, n\}$. Let $(u_1, u_2, \ldots, u_k) \in [1] \times [2] \times \cdots \times [k]$ and $(v_1, v_2, \ldots, v_k) \in [1] \times [2] \times \cdots \times [k]$ be two k-tuples satisfying

$$t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{k,u_k} = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{k,v_k}.$$

Then, $(u_1, u_2, \ldots, u_k) = (v_1, v_2, \ldots, v_k)$.

Proof of Lemma 7.79. We shall prove Lemma 7.79 by induction over k:

Induction base: Lemma 7.79 is true when $k = 0$ [371].

Induction step: Let $K \in \{0, 1, \ldots, n\}$ be positive. Assume that Lemma 7.79 holds when $k = K - 1$. We must show that Lemma 7.79 holds when $k = K$.

We have assumed that Lemma 7.79 holds when $k = K - 1$. In other words, the following statement holds:

[371] *Proof.* There exists only one 0-tuple: namely, the empty 0-tuple $()$. Thus, any two 0-tuples are equal.

Lemma 7.79 claims the equality of two k-tuples. When $k = 0$, any two k-tuples are equal (since any two 0-tuples are equal). Hence, Lemma 7.79 is true when $k = 0$.

Statement 1: Let $n \in \mathbb{N}$. Assume that $K-1 \in \{0,1,\ldots,n\}$. Let $(u_1,u_2,\ldots,u_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$ and $(v_1,v_2,\ldots,v_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$ be two $(K-1)$-tuples satisfying

$$t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}} = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}.$$

Then, $(u_1,u_2,\ldots,u_{K-1}) = (v_1,v_2,\ldots,v_{K-1})$.

We must show that Lemma 7.79 holds when $k=K$. In other words, we must prove the following statement:

Statement 2: Let $n \in \mathbb{N}$. Assume that $K \in \{0,1,\ldots,n\}$. Let $(u_1,u_2,\ldots,u_K) \in [1] \times [2] \times \cdots \times [K]$ and $(v_1,v_2,\ldots,v_K) \in [1] \times [2] \times \cdots \times [K]$ be two K-tuples satisfying

$$t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K,u_K} = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K,v_K}.$$

Then, $(u_1,u_2,\ldots,u_K) = (v_1,v_2,\ldots,v_K)$.

[*Proof of Statement 2:* We have $K \in \{0,1,\ldots,n\}$. Thus, $K \in \{1,2,\ldots,n\}$ (since K is positive), so that $K-1 \in \{0,1,\ldots,n-1\} \subseteq \{0,1,\ldots,n\}$.

We have $(u_1,u_2,\ldots,u_K) \in [1] \times [2] \times \cdots \times [K]$. In other words, $u_j \in [j]$ for each $j \in \{1,2,\ldots,K\}$. Hence, $(u_1,u_2,\ldots,u_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$. The same argument (applied to v_i instead of u_i) shows that $(v_1,v_2,\ldots,v_{K-1}) \in [1] \times [2] \times \cdots \times [K-1]$.

Define $\sigma \in S_n$ by $\sigma = t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K,u_K}$. Hence, Lemma 7.78 **(b)** (applied to $k=K$ and $i_j = u_j$) yields that $\sigma(u_K) = K$ (since $K > 0$). But we also have $\sigma = t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K,u_K} = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K,v_K}$. Hence, Lemma 7.78 **(b)** (applied to $k=K$ and $i_j = v_j$) yields that $\sigma(v_K) = K$ (since $K > 0$). Hence, $\sigma(u_K) = K = \sigma(v_K)$.

We have $\sigma \in S_n$. In other words, σ is a permutation of $[n]$ (since S_n is the set of all permutations of $[n]$). In other words, σ is a bijection $[n] \to [n]$. Hence, this map σ is bijective, thus injective. Therefore, from $\sigma(u_K) = \sigma(v_K)$, we conclude that $u_K = v_K$.

Recall that $u_j \in [j]$ for each $j \in \{1,2,\ldots,K\}$. Applying this to $j=K$, we obtain $u_K \in [K]$. In view of $u_K = v_K$, this rewrites as $v_K \in [K]$. But $K \leq n$ (since $K \in \{0,1,\ldots,n\}$), so that $[K] \subseteq [n]$. Hence, $v_K \in [K] \subseteq [n]$. Furthermore, $K \in \{1,2,\ldots,n\} = [n]$. Hence, Lemma 7.75 **(d)** (applied to K and v_K instead of i and j) yields $t_{K,v_K} \circ t_{K,v_K} = \mathrm{id}$.

Now, recall that $K > 0$. Hence,

$$\begin{aligned} & t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K,u_K} \\ & = \left(t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}}\right) \circ \underbrace{t_{K,u_K}}_{\substack{=t_{K,v_K} \\ (\text{since } u_K = v_K)}} = \left(t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}}\right) \circ t_{K,v_K}. \end{aligned}$$

Comparing this with

$$t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K,u_K}$$
$$= t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K,v_K} = \left(t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}\right) \circ t_{K,v_K},$$

we obtain

$$\left(t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}}\right) \circ t_{K,v_K} = \left(t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}\right) \circ t_{K,v_K}.$$

Thus,

$$\underbrace{\left(t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}}\right) \circ t_{K,v_K}}_{=\left(t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}\right) \circ t_{K,v_K}} \circ t_{K,v_K}$$
$$= \left(t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}\right) \circ \underbrace{t_{K,v_K} \circ t_{K,v_K}}_{=\mathrm{id}}$$
$$= t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}.$$

Comparing this with

$$\left(t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}}\right) \circ \underbrace{t_{K,v_K} \circ t_{K,v_K}}_{=\mathrm{id}} = t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}},$$

we obtain

$$t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{K-1,u_{K-1}} = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{K-1,v_{K-1}}.$$

Hence, Statement 1 shows that $(u_1, u_2, \ldots, u_{K-1}) = (v_1, v_2, \ldots, v_{K-1})$. In other words,

$$u_j = v_j \qquad \text{for each } j \in \{1, 2, \ldots, K-1\}.$$

Combining this with $u_K = v_K$, we conclude that

$$u_j = v_j \qquad \text{for each } j \in \{1, 2, \ldots, K\}.$$

In other words, $(u_1, u_2, \ldots, u_K) = (v_1, v_2, \ldots, v_K)$. This proves Statement 2.]

So Statement 2 is proven. In other words, Lemma 7.79 holds when $k = K$ (since Statement 2 is precisely the claim of Lemma 7.79 for $k = K$). This completes the induction step. Thus, Lemma 7.79 is proven. □

7.48.2. Solving Exercise 5.9

Solution to Exercise 5.9. We have $n \in \{0, 1, \ldots, n\}$ (since $n \in \mathbb{N}$).

We have ($\sigma(i) = i$ for each $i \in \{n+1, n+2, \ldots, n\}$). (Indeed, this is vacuously true, because there exists no $i \in \{n+1, n+2, \ldots, n\}$.) Hence, Lemma 7.77 (applied to $k = n$) yields that there exists a n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$.

Moreover, there exists at most one such n-tuple[372]. Hence, there is a unique n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$ (because there exists such an n-tuple, and there exists at most one such n-tuple). This solves Exercise 5.9. □

7.48.3. Some consequences

We shall now use the result of Exercise 5.9 to derive a few basic properties of symmetric groups. We begin with the following:

Corollary 7.80. Let $n \in \mathbb{N}$. The map

$$[1] \times [2] \times \cdots \times [n] \to S_n,$$
$$(i_1, i_2, \ldots, i_n) \mapsto t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$$

is well-defined and bijective.

Proof of Corollary 7.80. For each $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$, we have $t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n} \in S_n$ (since $t_{1,i_1}, t_{2,i_2}, \ldots, t_{n,i_n}$ are well-defined elements of S_n). Hence, the map

$$[1] \times [2] \times \cdots \times [n] \to S_n,$$
$$(i_1, i_2, \ldots, i_n) \mapsto t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$$

is well-defined. Let us denote this map by A.

---

[372] *Proof.* Let $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ be two such n-tuples. We shall prove that $(u_1, u_2, \ldots, u_n) = (v_1, v_2, \ldots, v_n)$.

We know that $(u_1, u_2, \ldots, u_n)$ is an n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$. In other words, $(u_1, u_2, \ldots, u_n)$ is an n-tuple in $[1] \times [2] \times \cdots \times [n]$ and satisfies $\sigma = t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{n,u_n}$. Thus, $t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{n,u_n} = \sigma$ and $(u_1, u_2, \ldots, u_n) \in [1] \times [2] \times \cdots \times [n]$ (since $(u_1, u_2, \ldots, u_n)$ is an n-tuple in $[1] \times [2] \times \cdots \times [n]$).

We know that $(v_1, v_2, \ldots, v_n)$ is an n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$. In other words, $(v_1, v_2, \ldots, v_n)$ is an n-tuple in $[1] \times [2] \times \cdots \times [n]$ and satisfies $\sigma = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{n,v_n}$. Thus, $(v_1, v_2, \ldots, v_n) \in [1] \times [2] \times \cdots \times [n]$ (since $(v_1, v_2, \ldots, v_n)$ is an n-tuple in $[1] \times [2] \times \cdots \times [n]$).

Now, $t_{1,u_1} \circ t_{2,u_2} \circ \cdots \circ t_{n,u_n} = \sigma = t_{1,v_1} \circ t_{2,v_2} \circ \cdots \circ t_{n,v_n}$. Hence, Lemma 7.79 (applied to $k = n$) yields $(u_1, u_2, \ldots, u_n) = (v_1, v_2, \ldots, v_n)$ (since $n \in \{0, 1, \ldots, n\}$).

Now, forget that we fixed $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$. We thus have proven that if $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ are two n-tuples $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$, then $(u_1, u_2, \ldots, u_n) = (v_1, v_2, \ldots, v_n)$. In other words, there exists at most one n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$. Qed.

The map A is injective[373] and surjective[374]. Hence, the map A is bijective. In other words, the map

$$\begin{aligned}[1] \times [2] \times \cdots \times [n] &\to S_n, \\ (i_1, i_2, \ldots, i_n) &\mapsto t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}\end{aligned}$$

is bijective (since this map is precisely the map A). This completes the proof of Corollary 7.80. □

We can use Corollary 7.80 to obtain the following:

Corollary 7.81. Let $n \in \mathbb{N}$. Then, $|S_n| = n!$.

Proof of Corollary 7.81. Corollary 7.80 shows that the map

$$\begin{aligned}[1] \times [2] \times \cdots \times [n] &\to S_n, \\ (i_1, i_2, \ldots, i_n) &\mapsto t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}\end{aligned}$$

---

[373] *Proof.* Let $\mathbf{x}$ and $\mathbf{y}$ be two elements of $[1] \times [2] \times \cdots \times [n]$ satisfying $A(\mathbf{x}) = A(\mathbf{y})$. We shall prove that $\mathbf{x} = \mathbf{y}$.

Define $\sigma \in S_n$ by $\sigma = A(\mathbf{x})$. Thus, $\sigma = A(\mathbf{x}) = A(\mathbf{y})$.

Write the element $\mathbf{x} \in [1] \times [2] \times \cdots \times [n]$ in the form $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Hence, $(x_1, x_2, \ldots, x_n) = \mathbf{x} \in [1] \times [2] \times \cdots \times [n]$. Now,

$$\sigma = A\left(\underbrace{\mathbf{x}}_{=(x_1,x_2,\ldots,x_n)}\right) = A((x_1, x_2, \ldots, x_n)) = t_{1,x_1} \circ t_{2,x_2} \circ \cdots \circ t_{n,x_n}$$

(by the definition of A). Thus, $(x_1, x_2, \ldots, x_n)$ is an n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$ (since $(x_1, x_2, \ldots, x_n) \in [1] \times [2] \times \cdots \times [n]$ and $\sigma = t_{1,x_1} \circ t_{2,x_2} \circ \cdots \circ t_{n,x_n}$). In other words, $\mathbf{x}$ is an n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$ (since $\mathbf{x} = (x_1, x_2, \ldots, x_n)$). Similarly, $\mathbf{y}$ is an n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$ (since $\sigma = A(\mathbf{y})$).

But Exercise 5.9 shows that there is a unique n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$. In particular, there exists **at most one** such n-tuple. In other words, if $\mathbf{u}$ and $\mathbf{v}$ are two n-tuples $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$, then $\mathbf{u} = \mathbf{v}$. Applying this to $\mathbf{u} = \mathbf{x}$ and $\mathbf{v} = \mathbf{y}$, we conclude that $\mathbf{x} = \mathbf{y}$ (since $\mathbf{x}$ and $\mathbf{y}$ are two n-tuples $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$).

Now, forget that we fixed $\mathbf{x}$ and $\mathbf{y}$. We thus have shown that if $\mathbf{x}$ and $\mathbf{y}$ are two elements of $[1] \times [2] \times \cdots \times [n]$ satisfying $A(\mathbf{x}) = A(\mathbf{y})$, then $\mathbf{x} = \mathbf{y}$. In other words, the map A is injective.

[374] *Proof.* Let $\sigma \in S_n$. Then, Exercise 5.9 shows that there is a unique n-tuple $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$ such that $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$. Consider this $(i_1, i_2, \ldots, i_n)$. The definition of A yields $A((i_1, i_2, \ldots, i_n)) = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n} = \sigma$. Thus, $\sigma = A\left(\underbrace{(i_1, i_2, \ldots, i_n)}_{\in [1] \times [2] \times \cdots \times [n]}\right) \in A([1] \times [2] \times \cdots \times [n])$.

Now, forget that we fixed σ. We thus have proven that each $\sigma \in S_n$ satisfies $\sigma \in A([1] \times [2] \times \cdots \times [n])$. In other words, $S_n \subseteq A([1] \times [2] \times \cdots \times [n])$. In other words, the map A is surjective.

is well-defined and bijective. Hence, this map is a bijection from $[1] \times [2] \times \cdots \times [n]$ to S_n. Thus, there exists a bijection from $[1] \times [2] \times \cdots \times [n]$ to S_n (namely, this map). Thus,

$$|S_n| = |[1] \times [2] \times \cdots \times [n]| = |[1]| \cdot |[2]| \cdot \cdots \cdot |[n]| = \prod_{k=1}^{n} \left| \underbrace{[k]}_{\substack{=\{1,2,\ldots,k\} \\ \text{(by the definition of } [k])}} \right|$$
$$= \prod_{k=1}^{n} \underbrace{|\{1,2,\ldots,k\}|}_{=k} = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot \cdots \cdot n = n!.$$

This proves Corollary 7.81. □

Corollary 7.81 can be generalized:

Corollary 7.82. Let X be a finite set. Then, the number of all permutations of X is $|X|!$.

To derive Corollary 7.82 from Corollary 7.81, we need a basic lemma:

Lemma 7.83. Let X and Y be two sets. Let $f : X \to Y$ be a bijection. Then, the map

$$\{\text{permutations of } X\} \to \{\text{permutations of } Y\},$$
$$\sigma \mapsto f \circ \sigma \circ f^{-1}$$

is well-defined and bijective.

Proof of Lemma 7.83. This is straightforward to check. (The inverse of this map is the map

$$\{\text{permutations of } Y\} \to \{\text{permutations of } X\},$$
$$\tau \mapsto f^{-1} \circ \tau \circ f.$$

) □

Proof of Corollary 7.82. Define $n \in \mathbb{N}$ by $n = |X|$. (This is well-defined, since X is a finite set.)

The definition of the symmetric group S_n shows that S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$. In other words,

$$S_n = \{\text{permutations of the set } \{1, 2, \ldots, n\}\}. \tag{749}$$

Define a finite set Y by $Y = \{1, 2, \ldots, n\}$. Then, $\left| \underbrace{Y}_{=\{1,2,\ldots,n\}} \right| = |\{1, 2, \ldots, n\}| = n$. Also,

$$\begin{aligned}
&\left\{\text{permutations of } \underbrace{Y}_{=\{1,2,\ldots,n\}}\right\} \\
&= \{\text{permutations of } \{1, 2, \ldots, n\}\} = \{\text{permutations of the set } \{1, 2, \ldots, n\}\} \\
&= S_n \qquad (\text{by (749)}) . \qquad (750)
\end{aligned}$$

But we have $|Y| = n = |X|$. In other words, the two finite sets X and Y have the same size. Hence, there exists a bijection $f : X \to Y$. Consider this f. Lemma 7.83 thus shows that the map

$$\{\text{permutations of } X\} \to \{\text{permutations of } Y\},$$
$$\sigma \mapsto f \circ \sigma \circ f^{-1}$$

is well-defined and bijective. Thus, this map is a bijection. Hence, there exists a bijection $\{\text{permutations of } X\} \to \{\text{permutations of } Y\}$ (namely, this map). Thus,

$$|\{\text{permutations of } X\}| = \left| \underbrace{\{\text{permutations of } Y\}}_{\substack{=S_n \\ (\text{by (750)})}} \right| = |S_n| = n!$$

(by Corollary 7.81). Now, the number of all permutations of X is $|\{\text{permutations of } X\}| = \underbrace{n}_{=|X|}! = |X|!$. This proves Corollary 7.82. □

7.49. Solution to Exercise 5.10

Solution to Exercise 5.10. **(a)** We shall prove that

$$\ell\left(t_{i,j}\right) = 2|j - i| - 1. \qquad (751)$$

[*Proof of (751):* We know that $t_{i,j}$ is the permutation in S_n which switches i with j while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged; on the other hand, $t_{j,i}$ is the permutation in S_n which switches j with i while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged. Comparing these two descriptions of $t_{i,j}$ and $t_{j,i}$, we immediately see that they are identical (since switching i with j is the same thing as switching j with i). Thus, $t_{i,j} = t_{j,i}$. Also, clearly, $|j - i| = |i - j|$. Hence, the claim (751) does not change if we switch i with j. Thus, we can WLOG assume that $i \leq j$ (because otherwise, we can just switch i with j). Assume this. Now, $i \leq j$, so that $i < j$ (since i and j are distinct). Hence, $j > i$, so that $j - i > 0$, so that $|j - i| = j - i$.

Set

$$A = \{(i,k) \mid k \in \{i+1, i+2, \ldots, j\}\} \qquad \text{and}$$
$$B = \{(k,j) \mid k \in \{i+1, i+2, \ldots, j-1\}\}.$$

These two sets A and B satisfy $|A| = j - i$ and $|B| = j - i - 1$. Also, it is easy to see that these sets A and B are disjoint[375]. Thus,

$$|A \cup B| = \underbrace{|A|}_{=j-i} + \underbrace{|B|}_{=j-i-1} = (j-i) + (j-i-1) = 2\underbrace{(j-i)}_{=|j-i|} - 1 = 2|j-i| - 1.$$

Now, let $\operatorname{Inv}(t_{i,j})$ denote the set of all inversions of $t_{i,j}$. Then, $\ell(t_{i,j}) = |\operatorname{Inv}(t_{i,j})|$ (because $\ell(t_{i,j})$ was defined as the number of inversions of $t_{i,j}$, which number is obviously $|\operatorname{Inv}(t_{i,j})|$).

We shall now show that $\operatorname{Inv}(t_{i,j}) = A \cup B$. Indeed, it is clearly enough to prove $A \cup B \subseteq \operatorname{Inv}(t_{i,j})$ and $\operatorname{Inv}(t_{i,j}) \subseteq A \cup B$. Proving that $A \cup B \subseteq \operatorname{Inv}(t_{i,j})$ means proving that every element of $A \cup B$ is an inversion of $t_{i,j}$; this is straightforward[376]. Proving that $\operatorname{Inv}(t_{i,j}) \subseteq A \cup B$ means proving that every inversion of $t_{i,j}$ belongs to $A \cup B$; this is equally straightforward (although more tiresome)[377]. Hence, both $A \cup B \subseteq \operatorname{Inv}(t_{i,j})$ and $\operatorname{Inv}(t_{i,j}) \subseteq A \cup B$ are proven, and we conclude that $\operatorname{Inv}(t_{i,j}) =$

---

[375] *Proof.* Every element of B is a pair (k,j) whose first entry is $> i$, whereas every element of A is a pair (i,k) whose first entry equals i. Thus, if the sets A and B had an element e in common, then e would be a pair whose first entry is $> i$ (since $e \in B$) and equals i (since $e \in A$) at the same time, which of course is impossible. Hence, A and B are disjoint.

[376] *Proof.* We want to show that $A \cup B \subseteq \operatorname{Inv}(t_{i,j})$. In other words, we want to prove that $e \in \operatorname{Inv}(t_{i,j})$ for every $e \in A \cup B$.

So let $e \in A \cup B$. Thus, either $e \in A$ or $e \in B$.

Let us first consider the case when $e \in A$. Thus, $e \in A = \{(i,k) \mid k \in \{i+1, i+2, \ldots, j\}\}$. In other words, e has the form $e = (i,k)$ for some $k \in \{i+1, i+2, \ldots, j\}$. Consider this k. The permutation $t_{i,j}$ switches i with j while leaving all other numbers fixed. Thus, the permutation $t_{i,j}$ leaves the numbers $i+1, i+2, \ldots, j-1$ fixed, while sending the number j to i. Consequently, $t_{i,j}$ sends the numbers $i+1, i+2, \ldots, j-1, j$ to $i+1, i+2, \ldots, j-1, i$, respectively. Notice that all of the latter numbers $i+1, i+2, \ldots, j-1, i$ are smaller than j. Thus, $t_{i,j}(p) < j$ for every $p \in \{i+1, i+2, \ldots, j\}$. Applying this to $p = k$, we conclude that $t_{i,j}(k) < j$.

We have $i < k$ (since $k \in \{i+1, i+2, \ldots, j\}$) but $t_{i,j}(i) = j > t_{i,j}(k)$ (since we have just showed that $t_{i,j}(k) < j$). Thus, (i,k) is an inversion of $t_{i,j}$. In other words, $(i,k) \in \operatorname{Inv}(t_{i,j})$. Thus, $e = (i,k) \in \operatorname{Inv}(t_{i,j})$.

Thus, we have proven that $e \in \operatorname{Inv}(t_{i,j})$ in the case when $e \in A$. A similar argument (but now using $t_{i,j}(k) > i$ instead of $t_{i,j}(k) < j$) shows that $e \in \operatorname{Inv}(t_{i,j})$ in the case when $e \in B$. Since either of these two cases must hold (because we have either $e \in A$ or $e \in B$), we thus conclude that $e \in \operatorname{Inv}(t_{i,j})$. This concludes the proof.

[377] *Proof.* We want to show that $\operatorname{Inv}(t_{i,j}) \subseteq A \cup B$. In other words, we want to prove that $c \in A \cup B$ for every $c \in \operatorname{Inv}(t_{i,j})$.

So let $c \in \operatorname{Inv}(t_{i,j})$. Thus, c is an inversion of $t_{i,j}$. In other words, c is a pair (u,v) of integers satisfying $1 \leq u < v \leq n$ and $t_{i,j}(u) > t_{i,j}(v)$. Consider this (u,v). Thus, $c = (u,v)$. Our goal is

$A \cup B$. Thus, $\left|\operatorname{Inv}\left(t_{i,j}\right)\right| = |A \cup B| = 2|j-i|-1$. Thus, $\ell\left(t_{i,j}\right) = \left|\operatorname{Inv}\left(t_{i,j}\right)\right| = 2|j-i|-1$. This proves (751).]

---

to show that $c \in A \cup B$.

The permutation $t_{i,j}$ switches i with j while leaving all other numbers fixed. It thus makes sense to analyze several cases separately, depending on which of the numbers u and v belongs to $\{i,j\}$. Four cases are possible:

Case 1: We have $u \in \{i,j\}$ and $v \in \{i,j\}$.

Case 2: We have $u \in \{i,j\}$ and $v \notin \{i,j\}$.

Case 3: We have $u \notin \{i,j\}$ and $v \in \{i,j\}$.

Case 4: We have $u \notin \{i,j\}$ and $v \notin \{i,j\}$.

Let us first consider Case 1. In this case, we have $u \in \{i,j\}$ and $v \in \{i,j\}$. Thus, u and v are two elements of $\{i,j\}$. Since $u < v$, this leaves only one possibility for the pair (u,v): namely, $(u,v) = (i,j)$. Thus,

$$c = (u,v) = (i,j) \in \{(i,k) \mid k \in \{i+1, i+2, \ldots, j\}\} \qquad (\text{since } j \in \{i+1, i+2, \ldots, j\}) \\ = A \subseteq A \cup B.$$

Thus, $c \in A \cup B$ is proven in Case 1.

Let us next consider Case 2. In this case, we have $u \in \{i,j\}$ and $v \notin \{i,j\}$. Since $v \notin \{i,j\}$, we have $t_{i,j}(v) = v$ (since $t_{i,j}$ leaves all numbers other than i and j unchanged). Thus, $t_{i,j}(u) > t_{i,j}(v) = v > u$ (since $u < v$). If we had $u = j$, then this would rewrite as $t_{i,j}(j) > j$, which would contradict $t_{i,j}(j) = i < j$. Thus, we cannot have $u = j$. Hence, we must have $u = i$ (since $u \in \{i,j\}$ forces u to be either i or j). But we have shown that $t_{i,j}(u) > v$, so that $v < t_{i,j}\left(\underbrace{u}_{=i}\right) = t_{i,j}(i) = j$ (since $t_{i,j}$ switches i with j). Combined with $v > u = i$, this yields $i < v < j$, so that $v \in \{i+1, i+2, \ldots, j-1\}$, so that

$$c = \left(\underbrace{u}_{=i}, v\right) = (i,v) \in \{(i,k) \mid k \in \{i+1, i+2, \ldots, j-1\}\} \\ (\text{since } v \in \{i+1, i+2, \ldots, j-1\}) \\ \subseteq \{(i,k) \mid k \in \{i+1, i+2, \ldots, j\}\} = A \subseteq A \cup B.$$

Thus, $c \in A \cup B$ is proven in Case 2.

Let us next consider Case 3. In this case, we have $u \notin \{i,j\}$ and $v \in \{i,j\}$. Since $u \notin \{i,j\}$, we have $t_{i,j}(u) = u$ (since $t_{i,j}$ leaves all numbers other than i and j unchanged). Thus, from $t_{i,j}(u) > t_{i,j}(v)$, we obtain $t_{i,j}(v) < t_{i,j}(u) = u < v$. If we had $v = i$, then this would rewrite as $t_{i,j}(i) < i$, which would contradict $t_{i,j}(i) = j > i$. Thus, we cannot have $v = i$. Hence, we must have $v = j$ (since $v \in \{i,j\}$ forces v to be either i or j). But we have shown that $t_{i,j}(v) < u$, so that $u > t_{i,j}\left(\underbrace{v}_{=j}\right) = t_{i,j}(j) = i$ (since $t_{i,j}$ switches i with j). Combined with $u < v = j$, this yields $i < u < j$, so that $u \in \{i+1, i+2, \ldots, j-1\}$, so that

$$c = \left(u, \underbrace{v}_{=j}\right) = (u,j) \in \{(k,j) \mid k \in \{i+1, i+2, \ldots, j-1\}\} \\ (\text{since } u \in \{i+1, i+2, \ldots, j-1\}) \\ = B \subseteq A \cup B.$$

(b) *First solution to Exercise 5.10* ***(b)****:* From (751), we have $\ell\left(t_{i,j}\right) = 2\left|j-i\right| - 1$.

But the integer $2\left|j-i\right| - 1$ is odd. Thus, $\left(-1\right)^{2\left|j-i\right|-1} = -1$. But the definition of $\left(-1\right)^{t_{i,j}}$ yields

$$\left(-1\right)^{t_{i,j}} = \left(-1\right)^{\ell\left(t_{i,j}\right)} = \left(-1\right)^{2\left|j-i\right|-1} \qquad \left(\text{since } \ell\left(t_{i,j}\right) = 2\left|j-i\right| - 1\right)$$
$$= -1.$$

This solves Exercise 5.10 **(b)**.

Second solution to Exercise 5.10 ***(b)****:* Here is an alternative solution of Exercise 5.10 **(b)** which makes no use of part **(a)**.

The set $\{1,2,\ldots,n\}$ has at least two distinct elements (namely, i and j). Hence, $n \geq 2$.

There exists a permutation $\sigma \in S_n$ such that $(i,j) = (\sigma(1), \sigma(2))$ [378]. Consider such a σ.

We have $(i,j) = (\sigma(1), \sigma(2))$, thus $(\sigma(1), \sigma(2)) = (i,j)$. In other words, $\sigma(1) = i$ and $\sigma(2) = j$.

We have $n \geq 2$. Hence, the permutation s_1 in S_n is well-defined. According to its definition, this permutation s_1 switches 1 with 2 but leaves all other numbers unchanged. In other words, we have $s_1(1) = 2$, $s_1(2) = 1$, and

$$s_1(k) = k \qquad \text{for every } k \in \{1,2,\ldots,n\} \text{ satisfying } k \notin \{1,2\}. \tag{752}$$

On the other hand, the permutation $t_{i,j}$ switches i with j while leaving all other elements of $\{1,2,\ldots,n\}$ unchanged. In other words, we have $t_{i,j}(i) = j$, $t_{i,j}(j) = i$ and

$$t_{i,j}(k) = k \qquad \text{for every } k \in \{1,2,\ldots,n\} \text{ satisfying } k \notin \{i,j\}. \tag{753}$$

Thus, $c \in A \cup B$ is proven in Case 3.

(Notice that Case 3 was very similar to Case 2 – almost like a mirror version of that case, if not for a slight asymmetry in our definition of the sets A and B.)

Let us finally consider Case 4. In this case, we have $u \notin \{i,j\}$ and $v \notin \{i,j\}$. Thus, $t_{i,j}(u) = u$ (as in Case 3) and $t_{i,j}(v) = v$ (as in Case 2), so that $t_{i,j}(u) = u < v = t_{i,j}(v)$. This contradicts $t_{i,j}(u) > t_{i,j}(v)$. This contradiction shows that Case 4 cannot happen; thus, we can ignore this case completely. (Or we can argue that because "ex falso quodlibet", we have $c \in A \cup B$ in Case 4.

[The principle of "ex falso quodlibet" (which we have already stated in Convention 2.37) says that from a false assertion, any arbitrary assertion follows. (For example, if $1 = 0$, then anything is true.) This is one of the basic principles in logic, and we could use it here to prove $c \in A \cup B$ in Case 4: Namely, since we have derived a contradiction (i.e., proven a false assertion) in Case 4, we see that any arbitrary assertion holds in Case 4; in particular, $c \in A \cup B$ holds in Case 4.])

We have now checked that $c \in A \cup B$ in each of the four cases 1, 2, 3 and 4 (or in each of the three cases 1, 2 and 3, and Case 4 never happens). Thus, $c \in A \cup B$ always holds. This completes our proof.

[378] *Proof.* Let $[n] = \{1,2,\ldots,n\}$. Notice that $n \geq 2$ and thus $2 \in \{0,1,\ldots,n\}$.

The integers i and j are distinct. Hence, (i,j) is a list of some elements of $[n]$ such that i and j are distinct. Therefore, Proposition 5.6 **(c)** (applied to $k = 2$ and $(p_1, p_2, \ldots, p_k) = (i,j)$) yields that there exists a permutation $\sigma \in S_n$ such that $(i,j) = (\sigma(1), \sigma(2))$. Qed.

Now, every $k \in \{1, 2, \ldots, n\}$ satisfies $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$ [379]. Hence, every $k \in \{1, 2, \ldots, n\}$ satisfies $(t_{i,j} \circ \sigma)(k) = t_{i,j}(\sigma(k)) = \sigma(s_1(k)) = (\sigma \circ s_1)(k)$. In other words, $t_{i,j} \circ \sigma = \sigma \circ s_1$.

Proposition 5.15 **(b)** shows that $(-1)^{s_k} = -1$ for every $k \in \{1, 2, \ldots, n-1\}$. Applied to $k = 1$, this yields $(-1)^{s_1} = -1$.

On the other hand, (315) (applied to $\tau = s_1$) yields $(-1)^{\sigma \circ s_1} = (-1)^{\sigma} \cdot \underbrace{(-1)^{s_1}}_{=-1} = (-1)^{\sigma} \cdot (-1)$.

But (315) (applied to $t_{i,j}$ and σ instead of σ and τ) yields $(-1)^{t_{i,j} \circ \sigma} = (-1)^{t_{i,j}} \cdot (-1)^{\sigma}$, so that

$$(-1)^{t_{i,j}} \cdot (-1)^{\sigma} = (-1)^{t_{i,j} \circ \sigma} = (-1)^{\sigma \circ s_1} \qquad \left(\text{since } t_{i,j} \circ \sigma = \sigma \circ s_1\right)$$
$$= (-1)^{\sigma} \cdot (-1).$$

We can cancel $(-1)^{\sigma}$ from this equality (since $(-1)^{\sigma} \in \{1, -1\}$ is a nonzero integer), and thus obtain $(-1)^{t_{i,j}} = -1$. This solves Exercise 5.10 **(b)** again. □

7.50. Solution to Exercise 5.11

Solution to Exercise 5.11. The inversions of w_0 are the pairs (i, j) of integers satisfying $1 \leq i < j \leq n$ and $w_0(i) > w_0(j)$ (because this is how we defined inversions). Since **every** pair of integers (i, j) satisfying $1 \leq i < j \leq n$ automatically satisfies

[379] *Proof.* Let $k \in \{1, 2, \ldots, n\}$. We need to show that $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$.

We are in one of the following three cases:

Case 1: We have $k = 1$.

Case 2: We have $k = 2$.

Case 3: We have $k \notin \{1, 2\}$.

Let us first consider Case 1. In this case, we have $k = 1$. Hence, $t_{i,j}\left(\sigma\left(\underbrace{k}_{=1}\right)\right) = t_{i,j}\left(\underbrace{\sigma(1)}_{=i}\right) = t_{i,j}(i) = j$. Compared with $\sigma\left(s_1\left(\underbrace{k}_{=1}\right)\right) = \sigma\left(\underbrace{s_1(1)}_{=2}\right) = \sigma(2) = j$, this yields $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$. Hence, $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$ is proven in Case 1.

The proof of $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$ in Case 2 is similar and left to the reader.

Let us first consider Case 3. In this case, we have $k \notin \{1, 2\}$. Hence, $s_1(k) = k$ (by (752)). On the other hand, the map σ is a permutation (since $\sigma \in S_n$), thus injective. Now, from $k \neq 1$ (since $k \notin \{1, 2\}$), we obtain $\sigma(k) \neq \sigma(1)$ (since the map σ is injective), so that $\sigma(k) \neq \sigma(1) = i$. Similarly, from $k \neq 2$, we can obtain $\sigma(k) \neq j$. Combining $\sigma(k) \neq i$ with $\sigma(k) \neq j$, we obtain $\sigma(k) \notin \{i, j\}$, and therefore $t_{i,j}(\sigma(k)) = \sigma(k)$ (by (753), applied to $\sigma(k)$ instead of k). Compared with $\sigma\left(\underbrace{s_1(k)}_{=k}\right) = \sigma(k)$, this yields $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$. Hence, $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$ is proven in Case 3.

Thus, $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$ is proven in each of the three Cases 1, 2 and 3. Hence, $t_{i,j}(\sigma(k)) = \sigma(s_1(k))$ always holds, qed.

$w_0(i) > w_0(j)$ [380], we can simplify this statement as follows: The inversions of w_0 are the pairs (i,j) of integers satisfying $1 \leq i < j \leq n$. But the number of such pairs is $n(n-1)/2$ [381]. Thus, the number of inversions of w_0 is $n(n-1)/2$. In other words, $\ell(w_0) = n(n-1)/2$ (since $\ell(w_0)$ is defined as the number of inversions of w_0). Therefore, the definition of $(-1)^{w_0}$ yields $(-1)^{w_0} = (-1)^{\ell(w_0)} = (-1)^{n(n-1)/2}$ (since $\ell(w_0) = n(n-1)/2$).

At this point, we could declare Exercise 5.11 to be solved, since we have found formulas for both $\ell(w_0)$ and $(-1)^{w_0}$. Nevertheless, let us give a different expression for $(-1)^{w_0}$, which can be evaluated faster. Namely, we claim that

$$(-1)^{w_0} = \begin{cases} 1, & \text{if } n \equiv 0 \bmod 4 \text{ or } n \equiv 1 \bmod 4; \\ -1, & \text{if } n \equiv 2 \bmod 4 \text{ or } n \equiv 3 \bmod 4 \end{cases}. \tag{754}$$

[*Proof of (754):* We must be in one of the following four cases:

Case 1: We have $n \equiv 0 \bmod 4$.

Case 2: We have $n \equiv 1 \bmod 4$.

Case 3: We have $n \equiv 2 \bmod 4$.

Case 4: We have $n \equiv 3 \bmod 4$.

The proofs of (754) in these four cases are more or less analogous. Let us only show the proof in Case 4. In this case, we have $n \equiv 3 \bmod 4$. Thus, $n = 4m + 3$ for

---

[380] *Proof.* Let (i,j) be a pair of integers satisfying $1 \leq i < j \leq n$. The definition of w_0 yields $w_0(i) = n+1-i$ and $w_0(j) = n+1-j$. Hence, $w_0(i) = n+1-\underbrace{i}_{<j} > n+1-j = w_0(j)$, qed.

[381] There are two ways to prove this:

- Either we can argue that these pairs are in a one-to-one correspondence with the 2-element subsets of $\{1,2,\ldots,n\}$. (Namely, any pair (i,j) corresponds to the subset $\{i,j\}$, and conversely, any subset S corresponds to the pair $(\min S, \max S)$.) Therefore, the number of such pairs equals the number of all 2-element subsets of $\{1,2,\ldots,n\}$; but the latter number is known to be $\dbinom{n}{2} = n(n-1)/2$.

- Alternatively, we can compute this number as follows: For every pair (i,j) of integers satisfying $1 \leq i < j \leq n$, we have $j \in \{2,3,\ldots,n\}$ (since $1 < j \leq n$). Hence,

$$\begin{aligned}
&(\text{the number of pairs } (i,j) \text{ of integers satisfying } 1 \leq i < j \leq n) \\
&= \sum_{k=2}^{n} \underbrace{(\text{the number of pairs } (i,j) \text{ satisfying } 1 \leq i < j \leq n \text{ and } j = k)}_{=(\text{the number of integers } i \text{ satisfying } 1 \leq i < k) = k-1} \\
&= \sum_{k=2}^{n} (k-1) = \sum_{j=1}^{n-1} j \qquad (\text{here, we substituted } j \text{ for } k-1 \text{ in the sum}) \\
&= \sum_{i=1}^{n-1} i = \frac{(n-1)((n-1)+1)}{2} \\
&\qquad (\text{by (14), applied to } n-1 \text{ instead of } n) \\
&= n(n-1)/2.
\end{aligned}$$

some $m \in \mathbb{Z}$. Consider this m. We have

$$\underbrace{n}_{=4m+3} \underbrace{(n-1)}_{\substack{=4m+2 \\ (\text{since } n=4m+3)}} /2 = (4m+3) \underbrace{(4m+2)/2}_{=2m+1} = (4m+3)(2m+1)$$
$$= 8m^2 + 10m + 3 = 2\left(4m^2 + 5m + 1\right) + 1.$$

Thus, the integer $n(n-1)/2$ is odd, so that $(-1)^{n(n-1)/2} = -1$. Hence,

$$(-1)^{w_0} = (-1)^{n(n-1)/2} = -1.$$

Compared with $\begin{cases} 1, & \text{if } n \equiv 0 \operatorname{mod} 4 \text{ or } n \equiv 1 \operatorname{mod} 4; \\ -1, & \text{if } n \equiv 2 \operatorname{mod} 4 \text{ or } n \equiv 3 \operatorname{mod} 4 \end{cases} = -1$ (since $n \equiv 3 \operatorname{mod} 4$), this yields $(-1)^{w_0} = \begin{cases} 1, & \text{if } n \equiv 0 \operatorname{mod} 4 \text{ or } n \equiv 1 \operatorname{mod} 4; \\ -1, & \text{if } n \equiv 2 \operatorname{mod} 4 \text{ or } n \equiv 3 \operatorname{mod} 4 \end{cases}$. Thus, (754) is proven in Case 4. The other three cases are analogous, and so we conclude that (754) holds.] □

7.51. Solution to Exercise 5.12

Solution to Exercise 5.12. **(a)** Let σ be a permutation of X. We need to prove that $(-1)^{\sigma}_{\phi}$ depends only on the permutation σ of X, but not on the bijection ϕ. In other words, we need to prove that any two different choices of ϕ will lead to the same $(-1)^{\sigma}_{\phi}$. In other words, we need to prove that if ϕ_1 and ϕ_2 are two bijections $\phi : X \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$ (possibly distinct), then $(-1)^{\sigma}_{\phi_1} = (-1)^{\sigma}_{\phi_2}$.

So let ϕ_1 and ϕ_2 be two bijections $\phi : X \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$ (possibly distinct). We must show that $(-1)^{\sigma}_{\phi_1} = (-1)^{\sigma}_{\phi_2}$.

The map σ is a permutation of X, thus a bijection $X \to X$.

We know that ϕ_1 is a bijection $X \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. In other words, there exists some $n \in \mathbb{N}$ such that ϕ_1 is a bijection $X \to \{1, 2, \ldots, n\}$. Denote this n by n_1. Thus, ϕ_1 is a bijection $X \to \{1, 2, \ldots, n_1\}$. The definition of $(-1)^{\sigma}_{\phi_1}$ yields $(-1)^{\sigma}_{\phi_1} = (-1)^{\phi_1 \circ \sigma \circ \phi_1^{-1}}$.

The map ϕ_1 is a bijection. Thus, its inverse ϕ_1^{-1} is well-defined and also a bijection.

The map $\phi_1 \circ \sigma \circ \phi_1^{-1} : \{1, 2, \ldots, n_1\} \to \{1, 2, \ldots, n_1\}$ is a bijection (since it is a composition of the three bijections ϕ_1, σ and ϕ_1^{-1}). In other words, the map $\phi_1 \circ \sigma \circ \phi_1^{-1}$ is a permutation of $\{1, 2, \ldots, n_1\}$. Thus, $\phi_1 \circ \sigma \circ \phi_1^{-1} \in S_{n_1}$.

We thus have shown that ϕ_1^{-1} is well-defined and a bijection, and found an $n_1 \in \mathbb{N}$ such that $\phi_1 \circ \sigma \circ \phi_1^{-1} \in S_{n_1}$. Similarly, we can show that ϕ_2^{-1} is well-defined and a bijection, and find an $n_2 \in \mathbb{N}$ such that $\phi_2 \circ \sigma \circ \phi_2^{-1} \in S_{n_2}$. Consider this n_2. (We shall soon see that $n_1 = n_2$.) We have $(-1)^\sigma_{\phi_2} = (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}}$ (by the definition of $(-1)^\sigma_{\phi_2}$).

There exists a bijection from X to $\{1,2,\ldots,n_1\}$ (namely, ϕ_1). Hence, $|X| = |\{1,2,\ldots,n_1\}| = n_1$. Similarly, $|X| = n_2$. Thus, $n_1 = |X| = n_2$. Thus, we can define an $n \in \mathbb{N}$ by $n = n_1 = n_2$. Consider this n.

The map $\phi_2 \circ \phi_1^{-1} : \{1,2,\ldots,n_1\} \to \{1,2,\ldots,n_2\}$ is a bijection (since it is a composition of two bijections). Since $n_1 = n$ and $n_2 = n$, this rewrites as follows: The map $\phi_2 \circ \phi_1^{-1} : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ is a bijection. In other words, the map $\phi_2 \circ \phi_1^{-1}$ is a permutation of $\{1,2,\ldots,n\}$. Thus, $\phi_2 \circ \phi_1^{-1} \in S_n$.

We have $\phi_1 \circ \sigma \circ \phi_1^{-1} \in S_{n_1} = S_n$ (since $n_1 = n$) and $\phi_2 \circ \sigma \circ \phi_2^{-1} \in S_{n_2} = S_n$ (since $n_2 = n$).

Now, (315) (applied to $\phi_2 \circ \phi_1^{-1}$ and $\phi_1 \circ \sigma \circ \phi_1^{-1}$ instead of σ and τ) yields

$$(-1)^{\phi_2\circ\phi_1^{-1}\circ\phi_1\circ\sigma\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\phi_1^{-1}} \cdot (-1)^{\phi_1\circ\sigma\circ\phi_1^{-1}}.$$

Since $\phi_2 \circ \underbrace{\phi_1^{-1} \circ \phi_1}_{=\mathrm{id}} \circ \sigma \circ \phi_1^{-1} = \phi_2 \circ \sigma \circ \phi_1^{-1}$, this rewrites as

$$(-1)^{\phi_2\circ\sigma\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\phi_1^{-1}} \cdot (-1)^{\phi_1\circ\sigma\circ\phi_1^{-1}}. \qquad (755)$$

On the other hand, (315) (applied to $\phi_2 \circ \sigma \circ \phi_2^{-1}$ and $\phi_2 \circ \phi_1^{-1}$ instead of σ and τ) yields

$$(-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}\circ\phi_2\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}} \cdot (-1)^{\phi_2\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\phi_1^{-1}} \cdot (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}}.$$

Since $\phi_2 \circ \sigma \circ \underbrace{\phi_2^{-1} \circ \phi_2}_{=\mathrm{id}} \circ \phi_1^{-1} = \phi_2 \circ \sigma \circ \phi_1^{-1}$, this rewrites as

$$(-1)^{\phi_2\circ\sigma\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\phi_1^{-1}} \cdot (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}}.$$

Comparing this with (755), we obtain

$$(-1)^{\phi_2\circ\phi_1^{-1}} \cdot (-1)^{\phi_1\circ\sigma\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\phi_1^{-1}} \cdot (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}}.$$

We can cancel $(-1)^{\phi_2\circ\phi_1^{-1}}$ from this equality (since $(-1)^{\phi_2\circ\phi_1^{-1}} \in \{1,-1\}$ is a nonzero integer), and thus obtain $(-1)^{\phi_1\circ\sigma\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}}$. Hence,

$$(-1)^\sigma_{\phi_1} = (-1)^{\phi_1\circ\sigma\circ\phi_1^{-1}} = (-1)^{\phi_2\circ\sigma\circ\phi_2^{-1}} = (-1)^\sigma_{\phi_2}.$$

As we know, this completes the solution of Exercise 5.12 **(a)**.

(b) Fix a bijection $\phi : X \to \{1,2,\ldots,n\}$ for some $n \in \mathbb{N}$. (Such a bijection always exists.) We shall denote the identity permutation of X by id_X, so as to distinguish it from the identity permutation of $\{1, 2, \ldots, n\}$ (which we keep denoting by id without a subscript). The definition of $(-1)^{\operatorname{id}_X}$ now yields

$$\begin{aligned}(-1)^{\operatorname{id}_X} &= (-1)_\phi^{\operatorname{id}_X} = (-1)^{\phi\circ\operatorname{id}_X\circ\phi^{-1}} \qquad \left(\text{by the definition of } (-1)_\phi^{\operatorname{id}_X}\right) \\ &= (-1)^{\operatorname{id}} \qquad \left(\text{since } \phi\circ\operatorname{id}_X\circ\phi^{-1} = \phi\circ\phi^{-1} = \operatorname{id}\right) \\ &= 1.\end{aligned}$$

In other words, $(-1)^{\operatorname{id}} = 1$ for the identity permutation $\operatorname{id} : X \to X$ of X. Exercise 5.12 **(b)** is thus solved.

(c) Let σ and τ be two permutations of X. Fix a bijection $\phi : X \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. (Such a bijection always exists.) The definition of $(-1)^\sigma$ yields

$$(-1)^\sigma = (-1)_\phi^\sigma = (-1)^{\phi\circ\sigma\circ\phi^{-1}} \qquad \left(\text{by the definition of } (-1)_\phi^\sigma\right). \tag{756}$$

The definition of $(-1)^\tau$ yields

$$(-1)^\tau = (-1)_\phi^\tau = (-1)^{\phi\circ\tau\circ\phi^{-1}} \qquad \left(\text{by the definition of } (-1)_\phi^\tau\right). \tag{757}$$

The maps $\phi\circ\sigma\circ\phi^{-1}$ and $\phi\circ\tau\circ\phi^{-1}$ are permutations in S_n. Therefore, (315) (applied to $\phi\circ\sigma\circ\phi^{-1}$ and $\phi\circ\tau\circ\phi^{-1}$ instead of σ and τ) yields

$$(-1)^{\phi\circ\sigma\circ\phi^{-1}\circ\phi\circ\tau\circ\phi^{-1}} = (-1)^{\phi\circ\sigma\circ\phi^{-1}} \cdot (-1)^{\phi\circ\tau\circ\phi^{-1}}.$$

Since $\phi\circ\sigma\circ\underbrace{\phi^{-1}\circ\phi}_{=\operatorname{id}}\circ\tau\circ\phi^{-1} = \phi\circ\sigma\circ\tau\circ\phi^{-1}$, this rewrites as

$$(-1)^{\phi\circ\sigma\circ\tau\circ\phi^{-1}} = (-1)^{\phi\circ\sigma\circ\phi^{-1}} \cdot (-1)^{\phi\circ\tau\circ\phi^{-1}}.$$

But the definition of $(-1)^{\sigma\circ\tau}$ yields

$$\begin{aligned}(-1)^{\sigma\circ\tau} &= (-1)_\phi^{\sigma\circ\tau} = (-1)^{\phi\circ\sigma\circ\tau\circ\phi^{-1}} \qquad \left(\text{by the definition of } (-1)_\phi^{\sigma\circ\tau}\right) \\ &= \underbrace{(-1)^{\phi\circ\sigma\circ\phi^{-1}}}_{\substack{=(-1)^\sigma \\ \text{(by (756))}}} \cdot \underbrace{(-1)^{\phi\circ\tau\circ\phi^{-1}}}_{\substack{=(-1)^\tau \\ \text{(by (757))}}} = (-1)^\sigma \cdot (-1)^\tau.\end{aligned}$$

This solves Exercise 5.12 **(c)**. □

7.52. Solution to Exercise 5.13

Solution to Exercise 5.13. **(b)** We start with some trivia on sets and subsets.

If A is any set, then $\mathcal{P}_2(A)$ shall denote the set of all 2-element subsets of A. In other words, $\mathcal{P}_2(A)$ is defined to be $\{S \subseteq A \mid |S| = 2\}$. For instance,

$$\mathcal{P}_2(\{3,6,7\}) = \{\{3,6\},\{3,7\},\{6,7\}\};$$
$$\mathcal{P}_2(\{1,2,3,4\}) = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\};$$
$$\mathcal{P}_2(\{3\}) = \varnothing;$$
$$\mathcal{P}_2(\varnothing) = \varnothing.$$

[382]

If A and B are two sets, and if $f : A \to B$ is an injective map, then we can define a map $f_* : \mathcal{P}_2(A) \to \mathcal{P}_2(B)$ by

$$(f_*(S) = f(S) \qquad \text{for every } S \in \mathcal{P}_2(A)).$$

[383] This construction has the following three basic properties:

1. If A is a set, then $(\mathrm{id}_A)_* = \mathrm{id}_{\mathcal{P}_2(A)}$.
2. If A, B and C are three sets, and if $f : A \to B$ and $g : B \to C$ are two injective maps, then $(g \circ f)_* = g_* \circ f_*$. (Of course, $g \circ f$ is injective here, so $(g \circ f)_*$ makes sense.)
3. If A and B are two sets, and if $f : A \to B$ is an invertible map, then f_* is invertible as well and satisfies $(f^{-1})_* = (f_*)^{-1}$.

(The first of these three properties is obvious; the second follows by observing that $(g \circ f)(S) = g(f(S))$ for every $S \in \mathcal{P}_2(A)$; the third can be proven using the second or directly.)

Let $[n]$ be the set $\{1,2,\ldots,n\}$. Recall that S_n is the set of all permutations of the set $\{1,2,\ldots,n\}$. In other words, S_n is the set of all permutations of the set $[n]$ (since $\{1,2,\ldots,n\} = [n]$).

---

[382] Several authors write $\binom{A}{2}$ for $\mathcal{P}_2(A)$. This notation looks like a binomial coefficient, but with A at the top. Of course, this notation is chosen for its suggestiveness: When A is finite, the set $\binom{A}{2}$ satisfies $\left|\binom{A}{2}\right| = \binom{|A|}{2}$.

[383] At this point, we need to check that this map f_* is well-defined. Before I do this, let me rewrite the definition of f_* in a more intuitive way: An element of $\mathcal{P}_2(A)$ is a 2-element subset $\{a, a'\}$ of A. The map f_* takes this subset to $\{f(a), f(a')\}$ (in other words, it applies f to each of its elements).

So why is the map f_* well-defined? It is supposed to send every $S \in \mathcal{P}_2(A)$ to $f(S)$. Thus, in order to prove that it is well-defined, we need to show that $f(S) \in \mathcal{P}_2(B)$ for every $S \in \mathcal{P}_2(A)$.

Let $S \in \mathcal{P}_2(A)$. Thus, the set S is a 2-element subset of A. The map f sends its two elements to two **distinct** elements of B (they are distinct because f is injective). In other words, the set $f(S)$ has 2 elements. Thus, $f(S)$ is a 2-element subset of B; in other words, $f(S) \in \mathcal{P}_2(B)$. This proves that the map f_* is well-defined.

Notice that we have used the injectivity of f in this argument.

We have $\sigma \in S_n$. In other words, σ is a permutation of the set $[n]$ (since S_n is the set of all such permutations). Thus, σ is a bijective map $[n] \to [n]$. In particular, the map $\sigma_* : \mathcal{P}_2([n]) \to \mathcal{P}_2([n])$ is well-defined.

Recall that if A and B are two sets, and if $f : A \to B$ is an invertible map, then f_* is invertible as well. Applying this to $A = [n]$, $B = [n]$ and $f = \sigma$, we conclude that σ_* is invertible (since σ is invertible). In other words, σ_* is bijective.

Let G be the subset
$$\left\{(u,v) \in [n]^2 \mid u < v\right\}$$
of $[n]^2$.

For example, if $n = 4$, then
$$G = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}.$$
Comparing this with
$$\mathcal{P}_2([4]) = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\},$$
we observe that the set $\mathcal{P}_2([n])$ is obtained from G by "replacing all parentheses by brackets" (i.e., replacing each $(i,j) \in G$ by $\{i,j\}$). This holds for all n, not just for $n = 4$. Let us make this observation somewhat more bulletproof: Each $(i,j) \in G$ satisfies $\{i,j\} \in \mathcal{P}_2([n])$ (because $(i,j) \in G$ leads to $i < j$, so that $i \neq j$, so that $\{i,j\}$ is a 2-element set, and thus $\{i,j\} \in \mathcal{P}_2([n])$). Thus, we can define a map $\rho : G \to \mathcal{P}_2([n])$ by setting
$$(\rho((i,j)) = \{i,j\} \qquad \text{for every } (i,j) \in G).$$
This map ρ is injective (indeed, we can reconstruct every $(i,j) \in G$ from its image $\rho((i,j)) = \{i,j\}$, because $(i,j) \in G$ entails $i < j$) and surjective (since every two-element subset S of $[n]$ has the form $\{i,j\}$ for some $(i,j) \in [n]^2$ satisfying $i < j$). Hence, the map ρ is bijective.

For every $S \in \mathcal{P}_2([n])$, we define an element a_S of $\mathbb{C}$ by
$$a_S = a_{(\min S, \max S)}.$$
(In other words, for every $S \in \mathcal{P}_2([n])$, we define an element a_S of $\mathbb{C}$ by $a_S = a_{(i,j)}$, where i and j are the two elements of S in increasing order.)

Let $\operatorname{Inv}(\sigma)$ be the set of inversions of σ. Then, $(-1)^{\sigma} = (-1)^{|\operatorname{Inv}(\sigma)|}$ [384]. Furthermore, $\operatorname{Inv}(\sigma) \subseteq G$ [385].

Now, we notice the following facts:

[384] *Proof.* The definition of $\ell(\sigma)$ shows that
$$\begin{aligned}\ell(\sigma) &= (\text{the number of inversions of } \sigma) = (\text{the number of elements of } \operatorname{Inv}(\sigma)) \\ &\qquad (\text{since } \operatorname{Inv}(\sigma) \text{ is the set of all inversions of } \sigma) \\ &= |\operatorname{Inv}(\sigma)|.\end{aligned}$$
But the definition of $(-1)^{\sigma}$ yields $(-1)^{\sigma} = (-1)^{\ell(\sigma)} = (-1)^{|\operatorname{Inv}(\sigma)|}$ (since $\ell(\sigma) = |\operatorname{Inv}(\sigma)|$), qed.

[385] *Proof.* Let $c \in \operatorname{Inv}(\sigma)$. Thus, c is an inversion of σ (since $\operatorname{Inv}(\sigma)$ is the set of inversions of σ). In

- For every $(i,j) \in G$, we have

$$a_{(i,j)} = a_{\rho((i,j))}. \quad (758)$$

[386]

- For every $(i,j) \in \operatorname{Inv}(\sigma)$, we have

$$a_{(\sigma(i),\sigma(j))} = -a_{\sigma(\rho((i,j)))}. \quad (759)$$

[387]

- For every $(i,j) \in G$ satisfying $(i,j) \notin \operatorname{Inv}(\sigma)$, we have

$$a_{(\sigma(i),\sigma(j))} = a_{\sigma(\rho((i,j)))}. \quad (760)$$

[388].

---

other words, c is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. Consider this (i,j). We have $i \in [n]$ (since $1 \leq i \leq n$) and $j \in [n]$ (since $1 \leq j \leq n$). Thus, $(i,j) \in [n]^2$. Thus, (i,j) is an element of $[n]^2$ and satisfies $i < j$. In other words, $(i,j) \in \left\{(u,v) \in [n]^2 \mid u < v\right\} = G$. Hence, $c = (i,j) \in G$.

Now, let us forget that we fixed c. We thus have proven that every $c \in \operatorname{Inv}(\sigma)$ satisfies $c \in G$. In other words, $\operatorname{Inv}(\sigma) \subseteq G$, qed.

[386] *Proof of (758):* Let $(i,j) \in G$. Thus, $(i,j) \in G = \left\{(u,v) \in [n]^2 \mid u < v\right\}$. In other words, $(i,j) \in [n]^2$ and $i < j$.

The definition of ρ shows that $\rho((i,j)) = \{i,j\}$. Since $i < j$, this shows that the elements of the set $\rho((i,j))$ listed in increasing order are i and j. Hence, $\min(\rho((i,j))) = i$ and $\max(\rho((i,j))) = j$.

Now, the definition of $a_{\rho((i,j))}$ shows that $a_{\rho((i,j))} = a_{(\min(\rho((i,j))),\max(\rho((i,j))))} = a_{(i,j)}$ (since $\min(\rho((i,j))) = i$ and $\max(\rho((i,j))) = j$). This proves (758).

[387] *Proof of (759):* Let $(i,j) \in \operatorname{Inv}(\sigma)$. Thus, (i,j) is an inversion of σ. In other words, (i,j) is a pair of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$.

The definition of ρ shows that $\rho((i,j)) = \{i,j\}$. Hence, $\sigma\left(\underbrace{\rho((i,j))}_{=\{i,j\}}\right) = \sigma(\{i,j\}) = \{\sigma(i), \sigma(j)\}$. Since $\sigma(i) > \sigma(j)$, this shows that the elements of the set $\sigma(\rho((i,j)))$ listed in increasing order are $\sigma(j)$ and $\sigma(i)$. Hence, $\min(\sigma(\rho((i,j)))) = \sigma(j)$ and $\max(\sigma(\rho((i,j)))) = \sigma(i)$.

Now, the definition of $a_{\sigma(\rho((i,j)))}$ shows that

$$\begin{aligned} a_{\sigma(\rho((i,j)))} &= a_{(\min(\sigma(\rho((i,j)))),\max(\sigma(\rho((i,j)))))} = a_{(\sigma(j),\sigma(i))} \\ &\qquad \left(\text{since } \min(\sigma(\rho((i,j)))) = \sigma(j) \text{ and } \max(\sigma(\rho((i,j)))) = \sigma(i)\right) \\ &= -a_{(\sigma(i),\sigma(j))} \qquad \left(\text{by (321), applied to } \sigma(i) \text{ and } \sigma(j) \text{ instead of } i \text{ and } j\right). \end{aligned}$$

Hence, $a_{(\sigma(i),\sigma(j))} = -a_{\sigma(\rho((i,j)))}$. This proves (759).

[388] *Proof of (760):* Let $(i,j) \in G$ be such that $(i,j) \notin \operatorname{Inv}(\sigma)$. Thus, (i,j) is an element of $[n]^2$ satisfying $i < j$ (since $(i,j) \in G = \left\{(u,v) \in [n]^2 \mid u < v\right\}$).

Now, recall that $G = \left\{ (u, v) \in [n]^2 \mid u < v \right\}$. Hence, the product sign $\prod\limits_{1 \leq i < j \leq n}$

If we had $\sigma(i) > \sigma(j)$, then (i, j) would be an inversion of σ (since (i, j) is a pair of integers satisfying $1 \leq i < j \leq n$), and thus would belong to $\operatorname{Inv}(\sigma)$; this would contradict $(i, j) \notin \operatorname{Inv}(\sigma)$. Hence, we cannot have $\sigma(i) > \sigma(j)$. Thus, we have $\sigma(i) \leq \sigma(j)$. Since $\sigma(i) \neq \sigma(j)$ (because $i \neq j$ and because σ is injective), this shows that $\sigma(i) < \sigma(j)$.

The definition of ρ shows that $\rho((i, j)) = \{i, j\}$. Hence, $\sigma\left(\underbrace{\rho((i, j))}_{=\{i,j\}}\right) = \sigma(\{i, j\}) = \{\sigma(i), \sigma(j)\}$. Since $\sigma(i) < \sigma(j)$, this shows that the elements of the set $\sigma(\rho((i, j)))$ listed in increasing order are $\sigma(i)$ and $\sigma(j)$. Hence, $\min(\sigma(\rho((i, j)))) = \sigma(i)$ and $\max(\sigma(\rho((i, j)))) = \sigma(j)$.

Now, the definition of $a_{\sigma(\rho((i,j)))}$ shows that

$$a_{\sigma(\rho((i,j)))} = a_{(\min(\sigma(\rho((i,j)))), \max(\sigma(\rho((i,j)))))} = a_{(\sigma(i), \sigma(j))}$$
$$\left(\text{since } \min(\sigma(\rho((i, j)))) = \sigma(i) \text{ and } \max(\sigma(\rho((i, j)))) = \sigma(j)\right).$$

This proves (760).

in $\prod_{1\leq i<j\leq n} a_{(\sigma(i),\sigma(j))}$ can be replaced by the product sign $\prod_{(i,j)\in G}$. Thus, we have

$$\begin{aligned}
&\underbrace{\prod_{1\leq i<j\leq n}}_{=\prod_{(i,j)\in G}} a_{(\sigma(i),\sigma(j))} \\
&= \prod_{(i,j)\in G} a_{(\sigma(i),\sigma(j))} = \left(\prod_{\substack{(i,j)\in G;\\ (i,j)\in \operatorname{Inv}(\sigma)}} \underbrace{a_{(\sigma(i),\sigma(j))}}_{\substack{=-a_{\sigma(\rho((i,j)))}\\ \text{(by (759))}}} \right) \cdot \left(\prod_{\substack{(i,j)\in G;\\ (i,j)\notin \operatorname{Inv}(\sigma)}} \underbrace{a_{(\sigma(i),\sigma(j))}}_{\substack{=a_{\sigma(\rho((i,j)))}\\ \text{(by (760))}}} \right) \\
&= \underbrace{\left(\prod_{\substack{(i,j)\in G;\\ (i,j)\in \operatorname{Inv}(\sigma)}} \left(-a_{\sigma(\rho((i,j)))}\right) \right)}_{=(-1)^{\left|\{(i,j)\in G \mid (i,j)\in \operatorname{Inv}(\sigma)\}\right|} \left(\prod_{\substack{(i,j)\in G;\\ (i,j)\in \operatorname{Inv}(\sigma)}} a_{\sigma(\rho((i,j)))} \right)} \cdot \left(\prod_{\substack{(i,j)\in G;\\ (i,j)\notin \operatorname{Inv}(\sigma)}} a_{\sigma(\rho((i,j)))} \right) \\
&= \underbrace{(-1)^{\left|\{(i,j)\in G \mid (i,j)\in \operatorname{Inv}(\sigma)\}\right|}}_{\substack{=(-1)^{|\operatorname{Inv}(\sigma)|} \\ \text{(since } \operatorname{Inv}(\sigma)\subseteq G\text{, and thus} \\ \{(i,j)\in G \mid (i,j)\in \operatorname{Inv}(\sigma)\} = \operatorname{Inv}(\sigma))}} \underbrace{\left(\prod_{\substack{(i,j)\in G;\\ (i,j)\in \operatorname{Inv}(\sigma)}} a_{\sigma(\rho((i,j)))} \right) \cdot \left(\prod_{\substack{(i,j)\in G;\\ (i,j)\notin \operatorname{Inv}(\sigma)}} a_{\sigma(\rho((i,j)))} \right)}_{=\prod_{(i,j)\in G} a_{\sigma(\rho((i,j)))}} \\
&= \underbrace{(-1)^{|\operatorname{Inv}(\sigma)|}}_{=(-1)^{\sigma}} \underbrace{\prod_{(i,j)\in G} a_{\sigma(\rho((i,j)))}}_{\substack{=\prod_{T\in\mathcal{P}_2([n])} a_{\sigma(T)} \\ \text{(here, we have substituted } T \\ \text{for } \rho((i,j)) \text{ in the product,} \\ \text{since the map } \rho : G\to\mathcal{P}_2([n]) \\ \text{is bijective)}}} = (-1)^{\sigma} \cdot \prod_{T\in\mathcal{P}_2([n])} a_{\sigma(T)}.
\end{aligned} \tag{761}$$

On the other hand,

$$\underbrace{\prod_{1\le i<j\le n}}_{=\prod_{(i,j)\in G}} \underbrace{a_{(i,j)}}_{\substack{=a_{\rho((i,j))}\\ \text{(by (758))}}} = \prod_{(i,j)\in G} a_{\rho((i,j))} = \prod_{S\in\mathcal{P}_2([n])} a_S$$

$$\left(\begin{array}{c}\text{here, we have substituted } S \text{ for } \rho((i,j)) \text{ in the product,}\\ \text{since the map } \rho : G \to \mathcal{P}_2([n]) \text{ is bijective}\end{array}\right)$$

$$= \prod_{T\in\mathcal{P}_2([n])} \underbrace{a_{\sigma_*(T)}}_{\substack{=a_{\sigma(T)}\\ \text{(since } \sigma_*(T)=\sigma(T)\\ \text{(by the definition of } \sigma_*))}}$$

$$\left(\begin{array}{c}\text{here, we have substituted } \sigma_*(T) \text{ for } S \text{ in the product,}\\ \text{since the map } \sigma_* : \mathcal{P}_2([n]) \to \mathcal{P}_2([n]) \text{ is bijective}\end{array}\right)$$

$$= \prod_{T\in\mathcal{P}_2([n])} a_{\sigma(T)}. \tag{762}$$

Hence, (761) becomes

$$\prod_{1\le i<j\le n} a_{(\sigma(i),\sigma(j))} = (-1)^\sigma \cdot \underbrace{\prod_{T\in\mathcal{P}_2([n])} a_{\sigma(T)}}_{\substack{=\prod_{1\le i<j\le n} a_{(i,j)}\\ \text{(by (762))}}} = (-1)^\sigma \cdot \prod_{1\le i<j\le n} a_{(i,j)}.$$

This solves Exercise 5.13 **(b)**.

(a) We have $x_j - x_i = -\left(x_i - x_j\right)$ for every $(i,j) \in \{1,2,\ldots,n\}^2$. Hence, we can apply Exercise 5.13 **(b)** to $a_{(i,j)} = x_i - x_j$. As a result, we obtain $\prod_{1\le i<j\le n} \left(x_{\sigma(i)} - x_{\sigma(j)}\right) = (-1)^\sigma \cdot \prod_{1\le i<j\le n} \left(x_i - x_j\right)$. This solves Exercise 5.13 **(a)**.

(c) Applying Exercise 5.13 **(a)** to $x_i = i$, we obtain $\prod_{1\le i<j\le n} (\sigma(i) - \sigma(j)) = (-1)^\sigma \cdot \prod_{1\le i<j\le n} (i-j)$. We can divide both sides of this equality by $\prod_{1\le i<j\le n} (i-j)$ (because $\prod_{1\le i<j\le n} (i-j)$ is a product of nonzero integers, and thus nonzero). As a result, we obtain $\dfrac{\prod_{1\le i<j\le n} (\sigma(i) - \sigma(j))}{\prod_{1\le i<j\le n} (i-j)} = (-1)^\sigma$. Thus,

$$(-1)^\sigma = \frac{\prod_{1\le i<j\le n} (\sigma(i) - \sigma(j))}{\prod_{1\le i<j\le n} (i-j)} = \prod_{1\le i<j\le n} \frac{\sigma(i) - \sigma(j)}{i-j}.$$

Thus, (320) is proven. This solves Exercise 5.13 **(c)**.

(d) See below. □

Exercise 5.13 **(d)** asks us to give an alternative solution to Exercise 5.2 **(b)**. Let us do this now:

Alternative solution to Exercise 5.2 ***(b)***. Let $n \in \mathbb{N}$. Let σ and τ be two permutations in S_n. We need to show that $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \mod 2$.

We are going to prove that $(-1)^{\sigma\circ\tau} = (-1)^{\sigma} \cdot (-1)^{\tau}$ first.

Exercise 5.13 **(a)** (applied to $x_i = i$) yields

$$\prod_{1\leq i<j\leq n} (\sigma(i) - \sigma(j)) = (-1)^{\sigma} \cdot \prod_{1\leq i<j\leq n} (i-j).$$

Exercise 5.13 **(a)** (applied to $\sigma(i)$ and τ instead of x_i and σ) yields

$$\begin{aligned}
\prod_{1\leq i<j\leq n} (\sigma(\tau(i)) - \sigma(\tau(j))) &= (-1)^{\tau} \cdot \underbrace{\prod_{1\leq i<j\leq n} (\sigma(i) - \sigma(j))}_{=(-1)^{\sigma}\cdot \prod_{1\leq i<j\leq n}(i-j)} \\
&= \underbrace{(-1)^{\tau} \cdot (-1)^{\sigma}}_{=(-1)^{\sigma}\cdot(-1)^{\tau}} \cdot \prod_{1\leq i<j\leq n} (i-j) \\
&= (-1)^{\sigma} \cdot (-1)^{\tau} \cdot \prod_{1\leq i<j\leq n} (i-j). \qquad (763)
\end{aligned}$$

Exercise 5.13 **(a)** (applied to $\sigma \circ \tau$ instead of σ) yields

$$\prod_{1\leq i<j\leq n} ((\sigma\circ\tau)(i) - (\sigma\circ\tau)(j)) = (-1)^{\sigma\circ\tau} \cdot \prod_{1\leq i<j\leq n} (i-j).$$

Thus,

$$\begin{aligned}
(-1)^{\sigma\circ\tau} \cdot \prod_{1\leq i<j\leq n} (i-j) &= \prod_{1\leq i<j\leq n} \left(\underbrace{(\sigma\circ\tau)(i)}_{=\sigma(\tau(i))} - \underbrace{(\sigma\circ\tau)(j)}_{=\sigma(\tau(j))} \right) \\
&= \prod_{1\leq i<j\leq n} (\sigma(\tau(i)) - \sigma(\tau(j))) \\
&= (-1)^{\sigma} \cdot (-1)^{\tau} \cdot \prod_{1\leq i<j\leq n} (i-j) \qquad \text{(by (763))}. \qquad (764)
\end{aligned}$$

But the integer $\prod_{1\leq i<j\leq n} (i-j)$ is nonzero (since it is a product of the nonzero integers $i-j$). Hence, we can divide both sides of the equality (764) by $\prod_{1\leq i<j\leq n} (i-j)$. We thus obtain $(-1)^{\sigma\circ\tau} = (-1)^{\sigma} \cdot (-1)^{\tau}$.

Now, the definition of $(-1)^{\sigma}$ yields $(-1)^{\sigma} = (-1)^{\ell(\sigma)}$. Also, the definition of $(-1)^{\tau}$ yields $(-1)^{\tau} = (-1)^{\ell(\tau)}$. Hence,

$$\underbrace{(-1)^{\sigma}}_{=(-1)^{\ell(\sigma)}} \cdot \underbrace{(-1)^{\tau}}_{=(-1)^{\ell(\tau)}} = (-1)^{\ell(\sigma)} \cdot (-1)^{\ell(\tau)} = (-1)^{\ell(\sigma)+\ell(\tau)}.$$

Finally, the definition of $(-1)^{\sigma\circ\tau}$ yields $(-1)^{\sigma\circ\tau} = (-1)^{\ell(\sigma\circ\tau)}$. Thus,

$$(-1)^{\ell(\sigma\circ\tau)} = (-1)^{\sigma\circ\tau} = (-1)^{\sigma} \cdot (-1)^{\tau} = (-1)^{\ell(\sigma)+\ell(\tau)}.$$

But it is obvious that if two integers u and v satisfy $(-1)^u = (-1)^v$, then $u \equiv v \bmod 2$ [389]. Applying this to $u = \ell(\sigma\circ\tau)$ and $v = \ell(\sigma) + \ell(\tau)$, we obtain $\ell(\sigma\circ\tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2$. Thus, Exercise 5.2 **(b)** is solved again. □

7.53. Solution to Exercise 5.14

Exercise 5.14 is an example of a combinatorial fact that one can easily convince oneself of (with some handwaving), but that is quite hard to prove in a formal, bulletproof way. Thus, the solution given below is going to be long, but we hope that the reader can avoid major parts of it by figuring them out independently.

Before we start solving Exercise 5.14, let us define some notations.

Definition 7.84. For every $n \in \mathbb{N}$, we let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Definition 7.85. For every $n \in \mathbb{N}$ and every n-tuple $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ of integers, we define the following notations:

- An *inversion* of $\mathbf{a}$ will mean a pair $(i,j) \in [n]^2$ satisfying $i < j$ and $a_i > a_j$.
- We denote by $\operatorname{Inv}(\mathbf{a})$ the set of all inversions of $\mathbf{a}$. Thus, $\operatorname{Inv}(\mathbf{a}) \subseteq [n]^2$. More precisely,

$$\begin{aligned}
\operatorname{Inv}(\mathbf{a}) &= (\text{the set of all inversions of } \mathbf{a}) \\
&= \left\{(i,j) \in [n]^2 \mid i < j \text{ and } a_i > a_j\right\} \qquad (765) \\
&\qquad \left(\begin{array}{c}\text{since the inversions of } \mathbf{a} \text{ are the pairs } (i,j) \in [n]^2 \\ \text{satisfying } i < j \text{ and } a_i > a_j \text{ (by the definition} \\ \text{of an "inversion")}\end{array}\right) \\
&= \left\{(u,v) \in [n]^2 \mid u < v \text{ and } a_u > a_v\right\} \qquad (766) \\
&\qquad (\text{here, we renamed the index } (i,j) \text{ as } (u,v)).
\end{aligned}$$

- We denote by $\ell(\mathbf{a})$ the number $|\operatorname{Inv}(\mathbf{a})|$. (This is well-defined because $\operatorname{Inv}(\mathbf{a})$ is finite (since $\operatorname{Inv}(\mathbf{a}) \subseteq [n]^2$).) Thus,

$$\begin{aligned}
\ell(\mathbf{a}) &= \left|\underbrace{\operatorname{Inv}(\mathbf{a})}_{=(\text{the set of all inversions of } \mathbf{a})}\right| = |(\text{the set of all inversions of } \mathbf{a})| \\
&= (\text{the number of all inversions of } \mathbf{a}). \qquad (767)
\end{aligned}$$

[389] Indeed, Proposition 2.159 shows that the two statements $(u \equiv v \bmod 2)$ and $\left((-1)^u = (-1)^v\right)$ are equivalent.

The next, nearly trivial, lemma relies on the notion of the "list of all elements of S in increasing order (with no repetitions)", where S is a finite set of integers. This notion means exactly what it says (but see Definition 2.50 for a rigorous definition).

Lemma 7.86. Let P be a finite set of integers. Let $m = |P|$. Let $\sigma \in S_m$. Let $(p_1, p_2, \ldots, p_m)$ be the list of all elements of P in increasing order (with no repetitions). Let $\operatorname{Inv}(\sigma)$ denote the set of all inversions of σ.

(a) We have $\operatorname{Inv}(\sigma) = \operatorname{Inv}\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right)$.

(b) We have $\ell(\sigma) = \ell\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right)$.

Proof of Lemma 7.86. The inversions of σ are the pairs (i, j) of integers satisfying $1 \leq i < j \leq m$ and $\sigma(i) > \sigma(j)$. Thus, $\operatorname{Inv}(\sigma)$ (which is the set of all inversions of σ) is the set of all such pairs (i, j). In other words,

$$\begin{aligned}
\operatorname{Inv}(\sigma) &= \left\{(i,j) \in \mathbb{Z}^2 \mid 1 \leq i < j \leq m \text{ and } \sigma(i) > \sigma(j)\right\} \\
&= \left\{(i,j) \in [m]^2 \mid i < j \text{ and } \sigma(i) > \sigma(j)\right\} \\
&\qquad \left(\begin{array}{c}\text{because the pairs } (i,j) \in \mathbb{Z}^2 \text{ satisfying } 1 \leq i < j \leq m \\ \text{are precisely the pairs } (i,j) \in [m]^2 \text{ satisfying } i < j\end{array}\right) \\
&= \left\{(u,v) \in [m]^2 \mid u < v \text{ and } \sigma(u) > \sigma(v)\right\} \qquad (768)
\end{aligned}$$

(here, we have renamed the index (i, j) as (u, v)).

But $(p_1, p_2, \ldots, p_m)$ is the list of all elements of P in increasing order (with no repetitions). Thus, $(p_1, p_2, \ldots, p_m)$ is a strictly increasing list. In other words, $p_1 < p_2 < \cdots < p_m$. Hence, if i and j are two elements of $[m]$, then we have the following logical equivalence:

$$\left(p_i > p_j\right) \iff (i > j). \qquad (769)$$

But (766) (applied to m, $\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right)$ and $p_{\sigma(i)}$ instead of n, $\mathbf{a}$ and a_i) yields

$$\begin{aligned}
&\operatorname{Inv}\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right) \\
&= \left\{(u,v) \in [m]^2 \mid u < v \text{ and } \underbrace{p_{\sigma(u)} > p_{\sigma(v)}}_{\substack{\text{this is equivalent to } \sigma(u)>\sigma(v) \\ \text{(by (769), applied to } i=\sigma(u) \text{ and } j=\sigma(v))}}\right\} \\
&= \left\{(u,v) \in [m]^2 \mid u < v \text{ and } \sigma(u) > \sigma(v)\right\} = \operatorname{Inv}(\sigma)
\end{aligned}$$

(by (768)). This proves Lemma 7.86 **(a)**.

(b) We know that $\ell(\sigma)$ is the number of all inversions of σ (by the definition of $\ell(\sigma)$). In other words, $\ell(\sigma)$ is the size of the set of all inversions of σ. In other words, $\ell(\sigma)$ is the size of $\operatorname{Inv}(\sigma)$ (since the set of all inversions of σ is $\operatorname{Inv}(\sigma)$). Hence,

$$\ell(\sigma) = \left| \underbrace{\operatorname{Inv}(\sigma)}_{\substack{=\operatorname{Inv}\left(p_{\sigma(1)},p_{\sigma(2)},\ldots,p_{\sigma(m)}\right) \\ \text{(by Lemma 7.86 (a))}}} \right| = \left| \operatorname{Inv}\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right) \right|$$
$$= \ell\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right)$$

(since $\ell\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right)$ is defined as $\left|\operatorname{Inv}\left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\right)\right|$). This proves Lemma 7.86 **(b)**. □

In the following, we shall use the Iverson bracket notation introduced in Definition 3.48.

Now, let us show some more lemmas:

Lemma 7.87. Let $n \in \mathbb{N}$. Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-tuple of integers. Then,

$$\ell(\mathbf{a}) = \sum_{i=1}^{n} \sum_{j=1}^{n} [i < j] \left[a_i > a_j\right].$$

Proof of Lemma 7.87. If $\mathcal{A}$ and $\mathcal{B}$ are two logical statements, then

$$[\mathcal{A}] [\mathcal{B}] = [\mathcal{A} \text{ and } \mathcal{B}] \tag{770}$$

(by Exercise 3.12 **(b)**).

We have

$$\begin{aligned}
&\underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n}}_{=\sum_{i\in[n]}\sum_{j\in[n]}=\sum_{(i,j)\in[n]^2}} \underbrace{[i<j]\left[a_i>a_j\right]}_{\substack{=\left[i<j \text{ and } a_i>a_j\right] \\ (\text{by (770)})}} \\
&= \sum_{(i,j)\in[n]^2} \left[i<j \text{ and } a_i>a_j\right] \\
&= \sum_{\substack{(i,j)\in[n]^2; \\ i<j \text{ and } a_i>a_j}} \underbrace{\left[i<j \text{ and } a_i>a_j\right]}_{\substack{=1 \\ (\text{since } i<j \text{ and } a_i>a_j)}} + \sum_{\substack{(i,j)\in[n]^2; \\ \text{not } \left(i<j \text{ and } a_i>a_j\right)}} \underbrace{\left[i<j \text{ and } a_i>a_j\right]}_{\substack{=0 \\ (\text{since not } \left(i<j \text{ and } a_i>a_j\right))}} \\
&= \sum_{\substack{(i,j)\in[n]^2; \\ i<j \text{ and } a_i>a_j}} 1 + \underbrace{\sum_{\substack{(i,j)\in[n]^2; \\ \text{not } \left(i<j \text{ and } a_i>a_j\right)}} 0}_{=0} = \sum_{\substack{(i,j)\in[n]^2; \\ i<j \text{ and } a_i>a_j}} 1 \\
&= \left|\underbrace{\left\{(i,j)\in[n]^2 \mid i<j \text{ and } a_i>a_j\right\}}_{\substack{=\operatorname{Inv}(\mathbf{a}) \\ (\text{by (765)})}}\right| \cdot 1 = \left|\operatorname{Inv}(\mathbf{a})\right| \cdot 1 \\
&= \left|\operatorname{Inv}(\mathbf{a})\right| = \ell(\mathbf{a}) \qquad (\text{since } \ell(\mathbf{a}) \text{ was defined to be } \left|\operatorname{Inv}(\mathbf{a})\right|).
\end{aligned}$$

This proves Lemma 7.87. □

Lemma 7.88. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-tuple of integers. Let $\mathbf{b} = (b_1, b_2, \ldots, b_m)$ be an m-tuple of integers. Let $\mathbf{c}$ be the $(n+m)$-tuple $(a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m)$ of integers. Then,

$$\ell(\mathbf{c}) = \ell(\mathbf{a}) + \ell(\mathbf{b}) + \sum_{(i,j)\in[n]\times[m]} \left[a_i > b_j\right].$$

Proof of Lemma 7.88. Lemma 7.87 yields

$$\ell(\mathbf{a}) = \sum_{i=1}^{n}\sum_{j=1}^{n} [i<j]\left[a_i>a_j\right]. \tag{771}$$

Also, Lemma 7.87 (applied to m, $\mathbf{b}$ and b_i instead of n, $\mathbf{a}$ and a_i) yields

$$\ell(\mathbf{b}) = \sum_{i=1}^{m}\sum_{j=1}^{m} [i<j]\left[b_i>b_j\right]. \tag{772}$$

Write the $(n+m)$-tuple $\mathbf{c}$ in the form $\mathbf{c} = (c_1, c_2, \ldots, c_{n+m})$. Thus,

$$(c_1, c_2, \ldots, c_{n+m}) = \mathbf{c} = (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m)$$

(by the definition of $\mathbf{c}$). In other words,

$$(c_i = a_i \qquad \text{for every } i \in \{1, 2, \ldots, n\}) \tag{773}$$

and

$$(c_i = b_{i-n} \qquad \text{for every } i \in \{n+1, n+2, \ldots, n+m\}) . \tag{774}$$

But Lemma 7.87 (applied to $n+m$, $\mathbf{c}$ and c_i instead of n, $\mathbf{a}$ and a_i) yields

$$\begin{aligned} \ell(\mathbf{c}) &= \sum_{i=1}^{n+m} \sum_{j=1}^{n+m} [i < j] [c_i > c_j] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n+m} [i < j] [c_i > c_j] + \sum_{i=n+1}^{n+m} \sum_{j=1}^{n+m} [i < j] [c_i > c_j] \end{aligned} \tag{775}$$

(since $0 \leq n \leq n+m$).

But every $i \in \{1, 2, \ldots, n\}$ satisfies

$$\sum_{j=1}^{n+m} [i < j] [c_i > c_j] = \sum_{j=1}^{n} [i < j] [a_i > a_j] + \sum_{j=1}^{m} [a_i > b_j] \tag{776}$$

[390].

[390] *Proof of (776):* Let $i \in \{1, 2, \ldots, n\}$. Thus, $c_i = a_i$ (by (773)).

For every $j \in \{1, 2, \ldots, n\}$, we have

$$c_j = a_j \tag{777}$$

(by (773), applied to j instead of i).

For every $j \in \{n+1, n+2, \ldots, n+m\}$, we have

$$c_j = b_{j-n} \tag{778}$$

(by (774), applied to j instead of i).

We have $i \leq n$ (since $i \in \{1, 2, \ldots, n\}$). For every $j \in \{n+1, n+2, \ldots, n+m\}$, we have $j \geq n+1 > n \geq i$ (since $i \leq n$), thus $i < j$, therefore

$$[i < j] = 1. \tag{779}$$

Recall that $0 \leq n \leq n+m$. Hence,

$$\begin{aligned} \sum_{j=1}^{n+m} [i < j] [c_i > c_j] &= \sum_{j=1}^{n} [i < j] \Bigg[\underbrace{c_i}_{=a_i} > \underbrace{c_j}_{\substack{=a_j \\ \text{(by (777))}}} \Bigg] + \sum_{j=n+1}^{n+m} \underbrace{[i < j]}_{\substack{=1 \\ \text{(by (779))}}} \Bigg[\underbrace{c_i}_{=a_i} > \underbrace{c_j}_{\substack{=b_{j-n} \\ \text{(by (778))}}} \Bigg] \\ &= \sum_{j=1}^{n} [i < j] [a_i > a_j] + \sum_{j=n+1}^{n+m} [a_i > b_{j-n}] \\ &= \sum_{j=1}^{n} [i < j] [a_i > a_j] + \sum_{j=1}^{m} [a_i > b_j] \\ &\qquad (\text{here, we have substituted } j \text{ for } j-n \text{ in the second sum}) . \end{aligned}$$

This proves (776).

Also, every $i \in \{n+1, n+2, \ldots, n+m\}$ satisfies

$$\sum_{j=1}^{n+m} [i < j] \left[c_i > c_j\right] = \sum_{j=1}^{m} [i - n < j] \left[b_{i-n} > b_j\right] \tag{780}$$

[391].

[391] *Proof of (780):* Let $i \in \{n+1, n+2, \ldots, n+m\}$. Thus, $c_i = b_{i-n}$ (by (774)).

For every $j \in \{n+1, n+2, \ldots, n+m\}$, we have

$$c_j = b_{j-n} \tag{781}$$

(by (774), applied to j instead of i).

We have $i \geq n+1$ (since $i \in \{n+1, n+2, \ldots, n+m\}$) and thus $i \geq n+1 > n$, so that $n < i$. For every $j \in \{1, 2, \ldots, n\}$, we have $j \leq n \leq i$ (since $i \geq n$) and thus $i \geq j$. Hence, for every $j \in \{1, 2, \ldots, n\}$, we don't have $i < j$. Thus, for every $j \in \{1, 2, \ldots, n\}$, we have

$$[i < j] = 0. \tag{782}$$

Recall that $0 \leq n \leq n+m$. Hence,

$$\begin{aligned}
\sum_{j=1}^{n+m} [i < j] \left[c_i > c_j\right] &= \sum_{j=1}^{n} \underbrace{[i < j]}_{\substack{=0 \\ \text{(by (782))}}} \left[c_i > c_j\right] + \sum_{j=n+1}^{n+m} \underbrace{[i < j]}_{\substack{=[i-n<j-n] \\ \text{(since } i<j \text{ is equivalent to} \\ i-n<j-n)}} \left[\underbrace{c_i}_{=b_{i-n}} > \underbrace{c_j}_{\substack{=b_{j-n} \\ \text{(by (781))}}}\right] \\
&= \underbrace{\sum_{j=1}^{n} 0 \left[c_i > c_j\right]}_{=0} + \sum_{j=n+1}^{n+m} [i - n < j - n] \left[b_{i-n} > b_{j-n}\right] \\
&= \sum_{j=n+1}^{n+m} [i - n < j - n] \left[b_{i-n} > b_{j-n}\right] \\
&= \sum_{j=1}^{m} [i - n < j] \left[b_{i-n} > b_j\right] \\
&\qquad \left(\text{here, we have substituted } j \text{ for } j - n \text{ in the sum}\right).
\end{aligned}$$

This proves (780).

Now, (775) becomes

$$\begin{aligned}
&\ell(\mathbf{c}) \\
&= \sum_{i=1}^{n} \underbrace{\sum_{j=1}^{n+m} [i<j] \left[c_i > c_j\right]}_{\substack{=\sum_{j=1}^{n}[i<j]\left[a_i>a_j\right]+\sum_{j=1}^{m}\left[a_i>b_j\right] \\ \text{(by (776))}}} + \sum_{i=n+1}^{n+m} \underbrace{\sum_{j=1}^{n+m} [i<j] \left[c_i > c_j\right]}_{\substack{=\sum_{j=1}^{m}[i-n<j]\left[b_{i-n}>b_j\right] \\ \text{(by (780))}}} \\
&= \underbrace{\sum_{i=1}^{n} \left(\sum_{j=1}^{n} [i<j] \left[a_i > a_j\right] + \sum_{j=1}^{m} \left[a_i > b_j\right] \right)}_{=\sum_{i=1}^{n}\sum_{j=1}^{n}[i<j]\left[a_i>a_j\right]+\sum_{i=1}^{n}\sum_{j=1}^{m}\left[a_i>b_j\right]} + \underbrace{\sum_{i=n+1}^{n+m} \sum_{j=1}^{m} [i-n<j] \left[b_{i-n} > b_j\right]}_{\substack{=\sum_{i=1}^{m}\sum_{j=1}^{m}[i<j]\left[b_i>b_j\right] \\ \text{(here, we have substituted } i \text{ for } i-n \\ \text{in the outer sum)}}} \\
&= \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{n} [i<j] \left[a_i > a_j\right]}_{\substack{=\ell(\mathbf{a}) \\ \text{(by (771))}}} + \underbrace{\sum_{i=1}^{n}}_{=\sum_{i\in[n]}} \underbrace{\sum_{j=1}^{m}}_{=\sum_{j\in[m]}} \left[a_i > b_j\right] + \underbrace{\sum_{i=1}^{m} \sum_{j=1}^{m} [i<j] \left[b_i > b_j\right]}_{\substack{=\ell(\mathbf{b}) \\ \text{(by (772))}}} \\
&= \ell(\mathbf{a}) + \underbrace{\sum_{i\in[n]} \sum_{j\in[m]}}_{=\sum_{(i,j)\in[n]\times[m]}} \left[a_i > b_j\right] + \ell(\mathbf{b}) \\
&= \ell(\mathbf{a}) + \sum_{(i,j)\in[n]\times[m]} \left[a_i > b_j\right] + \ell(\mathbf{b}) = \ell(\mathbf{a}) + \ell(\mathbf{b}) + \sum_{(i,j)\in[n]\times[m]} \left[a_i > b_j\right].
\end{aligned}$$

This proves Lemma 7.88. □

Lemma 7.89. Let S be a finite set of integers. Let $(c_1, c_2, \ldots, c_s)$ be a list of all elements of S (with no repetitions).

(a) Then, the map $[s] \to S,\ h \mapsto c_h$ is well-defined and a bijection.

(b) Let $\pi \in S_s$. Then, $\left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$ is a list of all elements of S (with no repetitions).

Proof of Lemma 7.89. Lemma 7.89 should really be intuitively obvious; let me merely sketch how to formalize the intuition.

We have the following basic fact:

> *Statement 1:* Let X and Y be two sets. Let $\phi : X \to Y$ be a map. Then, ϕ is a bijection if and only if each $i \in Y$ has exactly one preimage under ϕ.

Statement 1 is well-known and easy to prove; it is a rather useful (necessary and sufficient) criterion for proving the bijectivity of maps (particularly since it does not require proving injectivity and surjectivity separately).

We assumed that $(c_1, c_2, \ldots, c_s)$ is a list of all elements of S (with no repetitions). This means that the following two statements are valid:

> *Statement 2:* The list $(c_1, c_2, \ldots, c_s)$ is a list of elements of S. (In other words, we have $c_h \in S$ for every $h \in [s]$.)
>
> *Statement 3:* Each element of S appears exactly once in the list $(c_1, c_2, \ldots, c_s)$. In other words, for each $i \in S$, there exists exactly one $h \in [s]$ satisfying $i = c_h$.

Statement 2 shows that $c_h \in S$ for every $h \in [s]$. Thus, the map $[s] \to S,\ h \mapsto c_h$ is well-defined. Denote this map by α. Thus, $\alpha(h) = c_h$ for every $h \in [s]$.

(a) Statement 3 shows that, for each $i \in S$, there exists exactly one $h \in [s]$ satisfying $i = c_h$. Since $\alpha(h) = c_h$ for every $h \in [s]$, we can rewrite this as follows: For each $i \in S$, there exists exactly one $h \in [s]$ satisfying $i = \alpha(h)$. In other words, each $i \in S$ has exactly one preimage under α. According to Statement 1 (applied to $X = [s]$, $Y = S$ and $\phi = \alpha$), this holds if and only if α is a bijection. Thus, α is a bijection. In other words, the map $[s] \to S,\ h \mapsto c_h$ is a bijection (because α is precisely this map). Thus, Lemma 7.89 **(a)** is proven.

(b) We must prove that $\left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$ is a list of all elements of S (with no repetitions). This means proving the following two statements:

> *Statement 4:* The list $\left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$ is a list of elements of S. (In other words, we have $c_{\pi(h)} \in S$ for every $h \in [s]$.)
>
> *Statement 5:* Each element of S appears exactly once in the list $\left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$. In other words, for each $i \in S$, there exists exactly one $h \in [s]$ satisfying $i = c_{\pi(h)}$.

[*Proof of Statement 4:* Statement 2 shows that $c_h \in S$ for every $h \in [s]$. Applying this to $\pi(h)$ instead of h, we conclude that $c_{\pi(h)} \in S$ for every $h \in [s]$. This proves Statement 4.]

[*Proof of Statement 5:* The map $\alpha \circ \pi$ is a bijection (since it is the composition of the two bijections α and π). According to Statement 1 (applied to $X = [s]$, $Y = S$ and $\phi = \alpha \circ \pi$), this means that each $i \in S$ has exactly one preimage under $\alpha \circ \pi$. In other words, for each $i \in S$, there exists exactly one $h \in [s]$ satisfying $i = (\alpha \circ \pi)(h)$. Since every $h \in [s]$ satisfies

$$(\alpha \circ \pi)(h) = \alpha(\pi(h)) = c_{\pi(h)} \qquad \text{(by the definition of } \alpha\text{)},$$

we can rewrite this as follows: For each $i \in S$, there exists exactly one $h \in [s]$ satisfying $i = c_{\pi(h)}$. This proves Statement 5.]

Now, Statements 4 and 5 are proven; thus, $\left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$ is a list of all elements of S (with no repetitions). This proves Lemma 7.89 **(b)**. □

Lemma 7.90. Let I be a finite set of integers. Let $k = |I|$. Then,

$$\sum_{x\in I}\sum_{y\in I}[x>y] = 0+1+\cdots+(k-1).$$

Proof of Lemma 7.90. Let $(i_1, i_2, \ldots, i_k)$ be the list of all elements of I in increasing order (with no repetitions). (Such a list exists, since $|I| = k$.) Then, Lemma 7.89 **(a)** (applied to I, k and $(i_1, i_2, \ldots, i_k)$ instead of S, s and $(c_1, c_2, \ldots, c_s)$) shows that the map $[k] \to I, \ h \mapsto i_h$ is well-defined and a bijection.

We have $i_1 < i_2 < \cdots < i_k$ (because of how $(i_1, i_2, \ldots, i_k)$ was defined). Hence, if u and v are two elements of $[k]$, then $i_u > i_v$ holds if and only if $u > v$. In other words, if u and v are two elements of $[k]$, then

$$[i_u > i_v] = [u > v]. \tag{783}$$

Now, every $x \in I$ satisfies $\sum_{y\in I}[x>y] = \sum_{v\in[k]}[x>i_v]$ (here, we have substituted i_v for y in the sum, since the map $[k] \to I, \ h \mapsto i_h$ is a bijection). Thus,

$$\begin{aligned}
\sum_{x\in I}\underbrace{\sum_{y\in I}[x>y]}_{=\sum_{v\in[k]}[x>i_v]} &= \sum_{x\in I}\sum_{v\in[k]}[x>i_v] \\
&= \sum_{u\in[k]}\sum_{v\in[k]}\underbrace{[i_u>i_v]}_{\substack{=[u>v]\\ \text{(by (783))}}} \qquad \left(\begin{array}{c}\text{here, we have substituted } i_u\\ \text{for } x \text{ in the sum, since the}\\ \text{map } [k]\to I,\ h\mapsto i_h \text{ is a bijection}\end{array}\right) \\
&= \sum_{u\in[k]}\sum_{v\in[k]}[u>v].
\end{aligned} \tag{784}$$

But every $u \in [k]$ satisfies

$$\sum_{v\in[k]}[u>v] = u-1 \tag{785}$$

[392].

[392] *Proof of (785):* Let $u \in [k]$. Then,

$$\begin{aligned}
\underbrace{\sum_{v\in[k]}}_{=\sum_{v=1}^{k}}[u>v] &= \sum_{v=1}^{k}[u>v] = \sum_{v=1}^{u-1}\underbrace{[u>v]}_{\substack{=1\\ \text{(since } u>v\\ \text{(since } v\le u-1<u))}} + \sum_{v=u}^{k}\underbrace{[u>v]}_{\substack{=0\\ \text{(since we don't have } u>v\\ \text{(since } v\ge u))}} \qquad (\text{since } 1\le u\le k) \\
&= \sum_{v=1}^{u-1}1 + \underbrace{\sum_{v=u}^{k}0}_{=0} = \sum_{v=1}^{u-1}1 = (u-1)1 = u-1.
\end{aligned}$$

This proves (785).

Now, (784) becomes

$$\sum_{x\in I}\sum_{y\in I}[x>y] = \underbrace{\sum_{u\in[k]}}_{=\sum_{u=1}^{k}} \underbrace{\sum_{v\in[k]}[u>v]}_{\substack{=u-1 \\ \text{(by (785))}}} = \sum_{u=1}^{k}(u-1)$$

$$= \sum_{u=0}^{k-1} u \qquad \text{(here, we have substituted } u \text{ for } u-1 \text{ in the sum)}$$

$$= 0+1+\cdots+(k-1).$$

This proves Lemma 7.90. □

Lemma 7.91. Let S be a finite set of integers. Let $(c_1, c_2, \ldots, c_s)$ be a list of all elements of S (with no repetitions).

Let $p_1, p_2, \ldots, p_s$ be s pairwise distinct elements of S. Then, there exists a $\pi \in S_s$ such that $(p_1, p_2, \ldots, p_s) = \left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$.

Proof of Lemma 7.91. We know that $(c_1, c_2, \ldots, c_s)$ is a list of all elements of S. Hence, $\{c_1, c_2, \ldots, c_s\} = S$.

The elements $p_1, p_2, \ldots, p_s$ are pairwise distinct. In other words, if i and j are two distinct elements of $\{1, 2, \ldots, s\}$, then

$$p_i \neq p_j. \tag{786}$$

For every $i \in \{1, 2, \ldots, s\}$, there exists an $h \in \{1, 2, \ldots, s\}$ such that $p_i = c_h$ [393]. Fix such an h, and denote it by h_i. Thus, for every $i \in \{1, 2, \ldots, s\}$, we have defined an $h_i \in \{1, 2, \ldots, s\}$ such that

$$p_i = c_{h_i}. \tag{787}$$

Define a map $\varphi : \{1, 2, \ldots, s\} \to \{1, 2, \ldots, s\}$ by

$$(\varphi(i) = h_i \qquad \text{for every } i \in \{1, 2, \ldots, s\}).$$

Then, the map φ is injective[394] and therefore bijective[395]. Thus, φ is a bijective map $\{1, 2, \ldots, s\} \to \{1, 2, \ldots, s\}$. In other words, φ is a permutation of the set

[393] *Proof.* Let $i \in \{1, 2, \ldots, s\}$. Then, $p_i \in S$ (since $p_1, p_2, \ldots, p_s$ are s elements of S). Thus, $p_i \in S = \{c_1, c_2, \ldots, c_s\}$. In other words, there exists an $h \in \{1, 2, \ldots, s\}$ such that $p_i = c_h$. Qed.

[394] *Proof.* Let i and j be two elements of $\{1, 2, \ldots, s\}$ such that $\varphi(i) = \varphi(j)$. We shall prove that $i = j$.

The definition of φ yields $\varphi(i) = h_i$ and $\varphi(j) = h_j$. But (787) yields $p_i = c_{h_i}$. Also, (787) (applied to j instead of i) yields $p_j = c_{h_j}$. Now, $p_i = c_{h_i} = c_{h_j}$ (since $h_i = \varphi(i) = \varphi(j) = h_j$). Comparing this with $p_j = c_{h_j}$, we obtain $p_i = p_j$.

If the elements i and j were distinct, then we would have $p_i \neq p_j$ (by (786)); but this would contradict $p_i = p_j$. Hence, the elements i and j cannot be distinct. In other words, we have $i = j$.

Now, forget that we fixed i and j. We thus have proven that if i and j are two elements of $\{1, 2, \ldots, s\}$ such that $\varphi(i) = \varphi(j)$, then $i = j$. In other words, the map φ is injective. Qed.

[395] *Proof.* The set $\{1, 2, \ldots, s\}$ is finite and satisfies $|\{1, 2, \ldots, s\}| \geq |\{1, 2, \ldots, s\}|$. Hence, Lemma 1.5

$\{1,2,\ldots,s\}$. In other words, $\varphi \in S_s$ (since S_s is the set of all permutations of the set $\{1,2,\ldots,s\}$). Furthermore, $(p_1,p_2,\ldots,p_s) = \left(c_{\varphi(1)}, c_{\varphi(2)}, \ldots, c_{\varphi(s)}\right)$ [396]. Hence, φ is an element of S_s and satisfies $(p_1,p_2,\ldots,p_s) = \left(c_{\varphi(1)}, c_{\varphi(2)}, \ldots, c_{\varphi(s)}\right)$. Thus, there exists a $\pi \in S_s$ such that $(p_1,p_2,\ldots,p_s) = \left(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(s)}\right)$ (namely, $\pi = \varphi$). This proves Lemma 7.91. □

Lemma 7.92. Let $n \in \mathbb{N}$. Let I be a subset of $[n]$. Let $k = |I|$.

Let $(a_1,a_2,\ldots,a_k)$ be a list of all elements of I (with no repetitions). Let $(b_1,b_2,\ldots,b_{n-k})$ be a list of all elements of $[n] \setminus I$ (with no repetitions).

(a) There exists a unique $\sigma \in S_n$ satisfying

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) = (a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k}).$$

(b) Let $\sum I$ denote the sum of all elements of I. (Thus, $\sum I = \sum_{i \in I} i$.) Then,

$$\sum_{(i,j) \in [k] \times [n-k]} [a_i > b_j] = \sum I - (1 + 2 + \cdots + k).$$

Proof of Lemma 7.92. **(a)** The k elements $a_1,a_2,\ldots,a_k$ belong to I (since $(a_1,a_2,\ldots,a_k)$ is a list of all elements of I), and thus belong to $[n]$ as well (since $I \subseteq [n]$). Also, the k elements $a_1,a_2,\ldots,a_k$ are pairwise distinct (since $(a_1,a_2,\ldots,a_k)$ is a list with no repetitions).

The $n-k$ elements $b_1,b_2,\ldots,b_{n-k}$ belong to $[n] \setminus I$ (since $(b_1,b_2,\ldots,b_{n-k})$ is a list of all elements of $[n] \setminus I$), and thus belong to $[n]$ as well (since $[n] \setminus I \subseteq [n]$). Also, the $n-k$ elements $b_1,b_2,\ldots,b_{n-k}$ are pairwise distinct (since $(b_1,b_2,\ldots,b_{n-k})$ is a list with no repetitions).

The n elements $a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k}$ are pairwise distinct[397] and belong to

(applied to $U = \{1,2,\ldots,s\}$, $V = \{1,2,\ldots,s\}$ and $f = \varphi$) shows that we have the following logical equivalence:

$$(\varphi \text{ is injective}) \iff (\varphi \text{ is bijective}).$$

Hence, φ is bijective (since φ is injective). Qed.

[396] *Proof.* For every $i \in \{1,2,\ldots,s\}$, we have $\varphi(i) = h_i$ (by the definition of φ) and thus $c_{\varphi(i)} = c_{h_i} = p_i$ (by (787)). In other words, for every $i \in \{1,2,\ldots,s\}$, we have $p_i = c_{\varphi(i)}$. In other words, we have $(p_1,p_2,\ldots,p_s) = \left(c_{\varphi(1)}, c_{\varphi(2)}, \ldots, c_{\varphi(s)}\right)$. Qed.

[397] *Proof.* Assume the contrary. Thus, two of the n elements $a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k}$ are equal. These two elements are either two of the elements $a_1,a_2,\ldots,a_k$, or two of the elements $b_1,b_2,\ldots,b_{n-k}$, or one of the former and one of the latter (in some order). Hence, we are in one of the following three cases:

Case 1: Two of the k elements $a_1,a_2,\ldots,a_k$ are equal.

Case 2: Two of the $n-k$ elements $b_1,b_2,\ldots,b_{n-k}$ are equal.

Case 3: One of the k elements $a_1,a_2,\ldots,a_k$ is equal to one of the elements $b_1,b_2,\ldots,b_{n-k}$.

But Case 1 is impossible (since the k elements $a_1,a_2,\ldots,a_k$ are pairwise distinct), and Case 2 is

$[n]$ [398]. Furthermore, $(1,2,\ldots,n)$ is a list of all elements of $[n]$ (with no repetition). Hence, Lemma 7.91 (applied to $S=[n]$, $s=n$, $(c_1,c_2,\ldots,c_s)=(1,2,\ldots,n)$ and $(p_1,p_2,\ldots,p_s)=(a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k})$) yields that there exists a $\pi\in S_n$ such that

$$(a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k})=(\pi(1),\pi(2),\ldots,\pi(n)).$$

Consider this π.

We have $\pi\in S_n$ and $(\pi(1),\pi(2),\ldots,\pi(n))=(a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k})$.

But our goal is to show that there exists a unique $\sigma\in S_n$ satisfying

$$(\sigma(1),\sigma(2),\ldots,\sigma(n))=(a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k}). \tag{788}$$

We already know that there exists **at least one** such σ (namely, $\sigma=\pi$). Hence, it remains to show that there exists **at most one** such σ. In other words, it remains to show that if σ_1 and σ_2 are two elements $\sigma\in S_n$ satisfying (788), then $\sigma_1=\sigma_2$.

So let σ_1 and σ_2 be two elements $\sigma\in S_n$ satisfying (788). We must prove that $\sigma_1=\sigma_2$.

We know that σ_1 is an element $\sigma\in S_n$ satisfying (788). In other words, σ_1 is an element of S_n and satisfies

$$(\sigma_1(1),\sigma_1(2),\ldots,\sigma_1(n))=(a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k}). \tag{789}$$

Similarly, σ_2 is an element of S_n and satisfies

$$(\sigma_2(1),\sigma_2(2),\ldots,\sigma_2(n))=(a_1,a_2,\ldots,a_k,b_1,b_2,\ldots,b_{n-k}). \tag{790}$$

Comparing (789) with (790), we obtain

$$(\sigma_1(1),\sigma_1(2),\ldots,\sigma_1(n))=(\sigma_2(1),\sigma_2(2),\ldots,\sigma_2(n)).$$

In other words, $\sigma_1(i)=\sigma_2(i)$ for every $i\in[n]$. Since both σ_1 and σ_2 are maps $[n]\to[n]$, this shows that $\sigma_1=\sigma_2$. This is precisely what we wanted to show.

Thus, we have proven that there exists **at most one** $\sigma\in S_n$ satisfying (788). As explained above, this completes the proof of Lemma 7.92 **(a)**.

(b) Clearly,

$$\sum I=\sum_{i\in I}i=\sum_{x\in I}x \tag{791}$$

also impossible (since the $n-k$ elements $b_1,b_2,\ldots,b_{n-k}$ are pairwise distinct). Hence, we must be in Case 3. In other words, one of the k elements $a_1,a_2,\ldots,a_k$ is equal to one of the elements $b_1,b_2,\ldots,b_{n-k}$. In other words, we must have $a_i=b_j$ for some $i\in\{1,2,\ldots,k\}$ and some $j\in\{1,2,\ldots,n-k\}$. Consider these i and j. We have $a_i\in I$ (since the k elements $a_1,a_2,\ldots,a_k$ belong to I). But $b_j\in[n]\setminus I$ (since the $n-k$ elements $b_1,b_2,\ldots,b_{n-k}$ belong to $[n]\setminus I$) and thus $b_j\notin I$. Hence, $a_i=b_j\notin I$; but this contradicts $a_i\in I$. This contradiction proves that our assumption was wrong; qed.

[398] This is because the k elements $a_1,a_2,\ldots,a_k$ belong to $[n]$, and because the $n-k$ elements $b_1,b_2,\ldots,b_{n-k}$ belong to $[n]$.

(here, we have renamed the summation index i as x).

We know that $(a_1, a_2, \ldots, a_k)$ is a list of all elements of I (with no repetitions). Therefore, Lemma 7.89 **(a)** (applied to I, k and $(a_1, a_2, \ldots, a_k)$ instead of S, s and $(c_1, c_2, \ldots, c_s)$) shows that the map $[k] \to I,\ h \mapsto a_h$ is well-defined and a bijection.

We know that $(b_1, b_2, \ldots, b_{n-k})$ is a list of all elements of $[n] \setminus I$ (with no repetitions). Therefore, Lemma 7.89 **(a)** (applied to $[n] \setminus I$, $n-k$ and $(b_1, b_2, \ldots, b_{n-k})$ instead of S, s and $(c_1, c_2, \ldots, c_s)$) shows that the map $[n-k] \to [n] \setminus I,\ h \mapsto b_h$ is well-defined and a bijection.

Now,

$$\underbrace{\sum_{(i,j)\in[k]\times[n-k]}}_{=\sum_{i\in[k]}\sum_{j\in[n-k]}} [a_i > b_j] = \sum_{i\in[k]} \underbrace{\sum_{j\in[n-k]} [a_i > b_j]}_{\substack{=\sum_{y\in[n]\setminus I} [a_i > y] \\ \text{(here, we have substituted } y \text{ for } b_j \\ \text{in the sum, since the} \\ \text{map } [n-k]\to[n]\setminus I,\ h\mapsto b_h \text{ is a bijection)}}} = \sum_{i\in[k]} \sum_{y\in[n]\setminus I} [a_i > y]$$
$$= \sum_{x\in I} \sum_{y\in[n]\setminus I} [x > y] \tag{792}$$

(here, we have substituted x for a_i in the outer sum, since the map $[k] \to I,\ h \mapsto a_h$ is a bijection). But every $x \in I$ satisfies

$$\sum_{y\in[n]\setminus I} [x > y] = x - 1 - \sum_{y\in I} [x > y] \tag{793}$$

[399].

[399] *Proof of (793):* Let $x \in I$. Then, $x \in I \subseteq [n]$, so that $1 \leq x \leq n$.

Every $y \in [n]$ satisfies either $y \in I$ or $y \notin I$ (but not both). Hence,

$$\sum_{y\in[n]} [x > y] = \underbrace{\sum_{\substack{y\in[n];\\ y\in I}}}_{\substack{=\sum_{y\in I} \\ \text{(since } I\subseteq[n])}} [x > y] + \underbrace{\sum_{\substack{y\in[n];\\ y\notin I}}}_{=\sum_{y\in[n]\setminus I}} [x > y] = \sum_{y\in I} [x > y] + \sum_{y\in[n]\setminus I} [x > y].$$

Now, (792) becomes

$$\begin{aligned}
&\sum_{(i,j)\in[k]\times[n-k]} \left[a_i > b_j\right] \\
&= \sum_{x\in I} \underbrace{\sum_{y\in[n]\setminus I} [x>y]}_{\substack{=x-1-\sum_{y\in I}[x>y] \\ \text{(by (793))}}} = \sum_{x\in I}\left(x-1-\sum_{y\in I}[x>y]\right) \\
&= \underbrace{\sum_{x\in I} x}_{\substack{=\sum I \\ \text{(by (791))}}} - \underbrace{\sum_{x\in I} 1}_{=|I|\cdot 1=|I|=k} - \underbrace{\sum_{x\in I}\sum_{y\in I}[x>y]}_{\substack{=0+1+\cdots+(k-1) \\ \text{(by Lemma 7.90)}}} = \sum I - k - (0+1+\cdots+(k-1)) \\
&= \sum I - \underbrace{((0+1+\cdots+(k-1))+k)}_{=0+1+\cdots+k=1+2+\cdots+k} = \sum I - (1+2+\cdots+k).
\end{aligned}$$

This proves Lemma 7.92 **(b)**. ☐

Let us now come to the actual solution of Exercise 5.14:

Solution to Exercise 5.14. We have $I\subseteq\{1,2,\ldots,n\}=[n]$ (since $[n]=\{1,2,\ldots,n\}$). Hence, $|[n]\setminus I| = \Bigl|\underbrace{[n]}_{=\{1,2,\ldots,n\}}\Bigr| - \underbrace{|I|}_{=k} = \underbrace{|\{1,2,\ldots,n\}|}_{=n} - k = n-k$. Thus, $n-k=$

Comparing this with

$$\begin{aligned}
\underbrace{\sum_{y\in[n]}}_{=\sum_{y=1}^{n}} [x>y] &= \sum_{y=1}^{n}[x>y] \\
&= \sum_{y=1}^{x-1} \underbrace{[x>y]}_{\substack{=1 \\ \text{(since } x>y \\ \text{(since } y\leq x-1<x))}} + \sum_{y=x}^{n} \underbrace{[x>y]}_{\substack{=0 \\ \text{(since we don't have } x>y \\ \text{(since } y\geq x))}} \qquad (\text{since } 1\leq x\leq n) \\
&= \sum_{y=1}^{x-1} 1 + \underbrace{\sum_{y=x}^{n} 0}_{=0} = \sum_{y=1}^{x-1} 1 = (x-1)\,1 = x-1,
\end{aligned}$$

we obtain

$$\sum_{y\in I}[x>y] + \sum_{y\in[n]\setminus I}[x>y] = x-1.$$

Subtracting $\sum_{y\in I}[x>y]$ from both sides of this equality, we obtain $\sum_{y\in[n]\setminus I}[x>y] = x-1-\sum_{y\in I}[x>y]$. This proves (793).

$|[n] \setminus I| \in \mathbb{N}$ (since the cardinality of any finite set is $\in \mathbb{N}$). Hence, $n-k \geq 0$, so that $k \leq n$. Also, $k = |I| \in \mathbb{N}$.

We know that $(a_1, a_2, \ldots, a_k)$ is a list of all elements of I (with no repetitions). Thus, Lemma 7.89 **(b)** (applied to I, k, $(a_1, a_2, \ldots, a_k)$ and α instead of S, s, $(c_1, c_2, \ldots, c_s)$ and π) shows that $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ is a list of all elements of I (with no repetitions).

We know that $(b_1, b_2, \ldots, b_{n-k})$ is a list of all elements of $\{1, 2, \ldots, n\} \setminus I$ (with no repetitions). In other words, $(b_1, b_2, \ldots, b_{n-k})$ is a list of all elements of $[n] \setminus I$ (with no repetitions) (since $[n] = \{1, 2, \ldots, n\}$). Thus, Lemma 7.89 **(b)** (applied to $[n] \setminus I$, $n-k$, $(b_1, b_2, \ldots, b_{n-k})$ and β instead of S, s, $(c_1, c_2, \ldots, c_s)$ and π) shows that $\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ is a list of all elements of $[n] \setminus I$ (with no repetitions).

(a) Lemma 7.92 **(a)** (applied to $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ and $\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ instead of $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_{n-k})$) shows that there exists a unique $\sigma \in S_n$ satisfying

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right).$$

This solves Exercise 5.14 **(a)**.

(b) Let $\mathbf{a}$ be the k-tuple $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ of integers. Recall that $(a_1, a_2, \ldots, a_k)$ is the list of all elements of I in increasing order (with no repetitions). Moreover, $k = |I|$. Hence, Lemma 7.86 **(b)** (applied to I, k, α and $(a_1, a_2, \ldots, a_k)$ instead of P, m, σ and $(p_1, p_2, \ldots, p_m)$) yields

$$\ell(\alpha) = \ell\left(\underbrace{\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)}_{=\mathbf{a}}\right) = \ell(\mathbf{a}). \tag{794}$$

Let $\mathbf{b}$ be the $(n-k)$-tuple $\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ of integers. Recall that $(b_1, b_2, \ldots, b_{n-k})$ is the list of all elements of $\{1, 2, \ldots, n\} \setminus I$ in increasing order (with no repetitions). In other words, $(b_1, b_2, \ldots, b_{n-k})$ is the list of all elements of $[n] \setminus I$ in increasing order (with no repetitions) (since $[n] = \{1, 2, \ldots, n\}$). Moreover, $n-k = |[n] \setminus I|$ (since $|[n] \setminus I| = n-k$). Hence, Lemma 7.86 **(b)** (applied to $[n] \setminus I$, $n-k$, β and $(b_1, b_2, \ldots, b_{n-k})$ instead of P, m, σ and $(p_1, p_2, \ldots, p_m)$) yields

$$\ell(\beta) = \ell\left(\underbrace{\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)}_{=\mathbf{b}}\right) = \ell(\mathbf{b}). \tag{795}$$

Let $\mathbf{c}$ be the $(k+(n-k))$-tuple $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ of integers.

The permutation $\sigma_{I,\alpha,\beta}$ is the unique $\sigma \in S_n$ satisfying

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right). \qquad (796)$$

Thus, $\sigma_{I,\alpha,\beta}$ is a $\sigma \in S_n$ satisfying (796). In other words, we have

$$\begin{aligned}&\left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right)\\ &= \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) = \mathbf{c} \qquad (797)\end{aligned}$$

(since $\mathbf{c} = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$).

Now, $\{1,2,\ldots,n\}$ is a finite set of integers and satisfies $n = |\{1,2,\ldots,n\}|$. Moreover, $(1,2,\ldots,n)$ is the list of all elements of $\{1,2,\ldots,n\}$ in increasing order (with no repetitions). Hence, Lemma 7.86 **(b)** (applied to $\{1,2,\ldots,n\}$, n, $\sigma_{I,\alpha,\beta}$ and $(1,2,\ldots,n)$) yields

$$\begin{aligned}\ell\left(\sigma_{I,\alpha,\beta}\right) &= \ell\left(\underbrace{\left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right)}_{\substack{=\mathbf{c} \\ \text{(by (797))}}}\right)\\ &= \ell(\mathbf{c}) = \ell(\mathbf{a}) + \ell(\mathbf{b}) + \sum_{(i,j)\in[k]\times[n-k]} \left[a_{\alpha(i)} > b_{\beta(j)}\right] \qquad (798)\end{aligned}$$

(by Lemma 7.88 (applied to k, $n-k$, $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ and $\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ instead of n, m, $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_m)$)).

Now, recall that $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ is a list of all elements of I (with no repetitions). Also, recall that $\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ is a list of all elements of $[n] \setminus I$ (with no repetitions). Lemma 7.92 **(b)** (applied to $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ and $\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)$ instead of $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_{n-k})$) thus shows that

$$\sum_{(i,j)\in[k]\times[n-k]} \left[a_{\alpha(i)} > b_{\beta(j)}\right] = \sum I - (1+2+\cdots+k).$$

Hence, (798) becomes

$$\begin{aligned}\ell\left(\sigma_{I,\alpha,\beta}\right) &= \underbrace{\ell(\mathbf{a})}_{\substack{=\ell(\alpha) \\ \text{(by (794))}}} + \underbrace{\ell(\mathbf{b})}_{\substack{=\ell(\beta) \\ \text{(by (795))}}} + \underbrace{\sum_{(i,j)\in[k]\times[n-k]} \left[a_{\alpha(i)} > b_{\beta(j)}\right]}_{=\sum I-(1+2+\cdots+k)}\\ &= \ell(\alpha) + \ell(\beta) + \sum I - (1+2+\cdots+k).\end{aligned}$$

It thus remains to prove that $(-1)^{\sigma_{I,\alpha,\beta}} = (-1)^{\alpha} \cdot (-1)^{\beta} \cdot (-1)^{\sum I-(1+2+\cdots+k)}$.

But the definition of $(-1)^{\sigma_{I,\alpha,\beta}}$ yields

$$\begin{aligned}(-1)^{\sigma_{I,\alpha,\beta}} &= (-1)^{\ell\left(\sigma_{I,\alpha,\beta}\right)} = (-1)^{\ell(\alpha)+\ell(\beta)+\sum I-(1+2+\cdots+k)} \\ &\qquad \left(\text{since } \ell\left(\sigma_{I,\alpha,\beta}\right) = \ell(\alpha)+\ell(\beta)+\sum I-(1+2+\cdots+k)\right) \\ &= (-1)^{\ell(\alpha)} \cdot (-1)^{\ell(\beta)} \cdot (-1)^{\sum I-(1+2+\cdots+k)}.\end{aligned}$$

Comparing this with

$$\begin{aligned}&\underbrace{(-1)^{\alpha}}_{\substack{=(-1)^{\ell(\alpha)} \\ \text{(by the definition of } (-1)^{\alpha})}} \cdot \underbrace{(-1)^{\beta}}_{\substack{=(-1)^{\ell(\beta)} \\ \text{(by the definition of } (-1)^{\beta})}} \cdot (-1)^{\sum I-(1+2+\cdots+k)} \\ &= (-1)^{\ell(\alpha)} \cdot (-1)^{\ell(\beta)} \cdot (-1)^{\sum I-(1+2+\cdots+k)},\end{aligned}$$

this yields $(-1)^{\sigma_{I,\alpha,\beta}} = (-1)^{\alpha} \cdot (-1)^{\beta} \cdot (-1)^{\sum I-(1+2+\cdots+k)}$. This completes the solution of Exercise 5.14 **(b)**.

(c) For every $(\alpha,\beta) \in S_k \times S_{n-k}$, the permutation $\sigma_{I,\alpha,\beta} \in S_n$ satisfies

$$\begin{aligned}&\left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right) \\ &= \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) \qquad (799)\end{aligned}$$

[400].

For every $(\alpha,\beta) \in S_k \times S_{n-k}$, the element $\sigma_{I,\alpha,\beta}$ is a well-defined element of $\{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}$ [401]. Hence, the map

$$\begin{aligned}S_k \times S_{n-k} &\to \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}, \\ (\alpha,\beta) &\mapsto \sigma_{I,\alpha,\beta}\end{aligned}$$

is well-defined. Denote this map by μ.

[400] *Proof of (799):* Let $(\alpha,\beta) \in S_k \times S_{n-k}$. The permutation $\sigma_{I,\alpha,\beta}$ is the unique $\sigma \in S_n$ satisfying

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right). \qquad (800)$$

Thus, $\sigma_{I,\alpha,\beta}$ is a $\sigma \in S_n$ satisfying (800). In other words, we have

$$\left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right) = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right).$$

Qed.

[401] *Proof.* Let $(\alpha,\beta) \in S_k \times S_{n-k}$. We must show that the element $\sigma_{I,\alpha,\beta}$ is a well-defined element of $\{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}$.

Now, it is easy to see that the map μ is injective[402].

Our next goal is to show that the map μ is surjective.

Now,

$$\begin{aligned}
&\left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(k)\right) \\
&= \left(\text{the list of the first } k \text{ entries of the list } \underbrace{\left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right)}_{\substack{=\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) \\ \text{(by (799))}}}\right) \\
&= \left(\text{the list of the first } k \text{ entries of the list } \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right)\right) \\
&= \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right).
\end{aligned}$$

Hence,

$$\left\{\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(k)\right\} = \left\{a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right\} = I$$

(since $\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right)$ is a list of all elements of I). Now,

$$\sigma_{I,\alpha,\beta}(\{1,2,\ldots,k\}) = \left\{\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(k)\right\} = I.$$

So we know that $\sigma_{I,\alpha,\beta}$ is an element of S_n and satisfies $\sigma_{I,\alpha,\beta}(\{1,2,\ldots,k\}) = I$. In other words,

$$\sigma_{I,\alpha,\beta} \in \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}.$$

We thus have proven that $\sigma_{I,\alpha,\beta}$ is a well-defined element of $\{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}$. Qed.

[402] *Proof.* Let (α,β) and (α',β') be two elements of $S_k \times S_{n-k}$ satisfying $\mu(\alpha,\beta) = \mu(\alpha',\beta')$. We will show that $(\alpha,\beta) = (\alpha',\beta')$.

The definition of $\mu(\alpha,\beta)$ yields $\mu(\alpha,\beta) = \sigma_{I,\alpha,\beta}$. The definition of $\mu(\alpha',\beta')$ yields $\mu(\alpha',\beta') = \sigma_{I,\alpha',\beta'}$. Hence, $\sigma_{I,\alpha',\beta'} = \mu(\alpha',\beta') = \mu(\alpha,\beta) = \sigma_{I,\alpha,\beta}$.

Now, (799) (applied to (α',β') instead of (α,β)) yields

$$\begin{aligned}
&\left(\sigma_{I,\alpha',\beta'}(1), \sigma_{I,\alpha',\beta'}(2), \ldots, \sigma_{I,\alpha',\beta'}(n)\right) \\
&= \left(a_{\alpha'(1)}, a_{\alpha'(2)}, \ldots, a_{\alpha'(k)}, b_{\beta'(1)}, b_{\beta'(2)}, \ldots, b_{\beta'(n-k)}\right).
\end{aligned}$$

Comparing this with

$$\begin{aligned}
&\left(\sigma_{I,\alpha',\beta'}(1), \sigma_{I,\alpha',\beta'}(2), \ldots, \sigma_{I,\alpha',\beta'}(n)\right) \\
&= \left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right) \qquad \left(\text{since } \sigma_{I,\alpha',\beta'} = \sigma_{I,\alpha,\beta}\right) \\
&= \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right),
\end{aligned}$$

we obtain

$$\begin{aligned}
&\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) \\
&= \left(a_{\alpha'(1)}, a_{\alpha'(2)}, \ldots, a_{\alpha'(k)}, b_{\beta'(1)}, b_{\beta'(2)}, \ldots, b_{\beta'(n-k)}\right).
\end{aligned}$$

In fact, let $\gamma \in \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}$. We shall construct an $(\alpha,\beta) \in S_k \times S_{n-k}$ such that $\mu(\alpha,\beta) = \gamma$.

We have $\gamma \in \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}$. In other words, γ is an element of S_n and satisfies $\gamma(\{1,2,\ldots,k\}) = I$.

We have $\gamma \in S_n$. In other words, γ is a permutation of $\{1,2,\ldots,n\}$. Thus, γ is a bijective map $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$, and therefore also an injective map.

Every $i \in \{1,2,\ldots,k\}$ satisfies $\gamma(i) \in I$ [403]. Thus, $\gamma(1), \gamma(2), \ldots, \gamma(k)$ are k elements of I.

If u and v are two distinct elements of $\{1,2,\ldots,k\}$, then $\gamma(u) \neq \gamma(v)$ (since the map γ is injective). In other words, the k elements $\gamma(1), \gamma(2), \ldots, \gamma(k)$ are pairwise distinct.

On the other hand, recall that $(a_1, a_2, \ldots, a_k)$ is a list of all elements of I (with no repetitions). Thus, Lemma 7.91 (applied to I, k, $(a_1, a_2, \ldots, a_k)$ and $(\gamma(1), \gamma(2), \ldots, \gamma(k))$ instead of S, s, $(c_1, c_2, \ldots, c_s)$ and $(p_1, p_2, \ldots, p_s)$) yields that there exists a $\pi \in S_k$ such that $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = \left(a_{\pi(1)}, a_{\pi(2)}, \ldots, a_{\pi(k)}\right)$. Denote this π by α. Thus, α is a $\pi \in S_k$ such that $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = \left(a_{\pi(1)}, a_{\pi(2)}, \ldots, a_{\pi(k)}\right)$. In other words, α is an element of S_k and satisfies

$$(\gamma(1), \gamma(2), \ldots, \gamma(k)) = \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right). \tag{803}$$

Every $j \in \{1,2,\ldots,n-k\}$ satisfies $\gamma(k+j) \in \{1,2,\ldots,n\} \setminus I$ [404]. Thus,

In other words,

$$\left(a_{\alpha(i)} = a_{\alpha'(i)} \text{ for every } i \in \{1,2,\ldots,k\}\right) \tag{801}$$

and

$$\left(b_{\beta(j)} = b_{\beta'(j)} \text{ for every } j \in \{1,2,\ldots,n-k\}\right). \tag{802}$$

Now, we shall prove that $\alpha = \alpha'$.

Indeed, fix $i \in \{1,2,\ldots,k\}$. The list $(a_1, a_2, \ldots, a_k)$ has no repetitions. In other words, the elements $a_1, a_2, \ldots, a_k$ are pairwise distinct. Thus, if u and v are two elements of $\{1,2,\ldots,k\}$ such that $a_u = a_v$, then $u = v$. Applying this to $u = \alpha(i)$ and $v = \alpha'(i)$, we obtain $\alpha(i) = \alpha'(i)$ (since (801) yields $a_{\alpha(i)} = a_{\alpha'(i)}$).

Now, forget that we fixed i. We thus have proven that $\alpha(i) = \alpha'(i)$ for every $i \in \{1,2,\ldots,k\}$. Thus, $\alpha = \alpha'$ (since α and α' are permutations of $\{1,2,\ldots,k\}$).

We have thus proven $\alpha = \alpha'$ using the equalities (801). Similarly, we can prove $\beta = \beta'$ using the equalities (802).

Now, $\left(\underbrace{\alpha}_{=\alpha'}, \underbrace{\beta}_{=\beta'}\right) = (\alpha', \beta')$.

Now, forget that we fixed (α,β) and (α',β'). We thus have proven that if (α,β) and (α',β') are two elements of $S_k \times S_{n-k}$ satisfying $\mu(\alpha,\beta) = \mu(\alpha',\beta')$, then $(\alpha,\beta) = (\alpha',\beta')$. In other words, the map μ is injective. Qed.

[403] *Proof.* Let $i \in \{1,2,\ldots,k\}$. Then, $\gamma\left(\underbrace{i}_{\in\{1,2,\ldots,k\}}\right) \in \gamma(\{1,2,\ldots,k\}) = I$, qed.

[404] *Proof.* Let $j \in \{1,2,\ldots,n-k\}$. Then, $k+j \in \{k+1, k+2, \ldots, n\} \subseteq \{1,2,\ldots,n\}$. Hence, $\gamma(k+j)$

$\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)$ are $n-k$ elements of $\{1,2,\ldots,n\} \setminus I$.

If u and v are two distinct elements of $\{k+1, k+2, \ldots, n\}$, then $\gamma(u) \neq \gamma(v)$ (since the map γ is injective). In other words, the $n-k$ elements $\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)$ are pairwise distinct.

On the other hand, recall that $(b_1, b_2, \ldots, b_{n-k})$ is a list of all elements of $\{1,2,\ldots,n\} \setminus I$ (with no repetitions). Thus, Lemma 7.91 (applied to $\{1,2,\ldots,n\} \setminus I$, $n-k$, $(b_1, b_2, \ldots, b_{n-k})$ and $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n))$ instead of S, s, $(c_1, c_2, \ldots, c_s)$ and $(p_1, p_2, \ldots, p_s)$) yields that there exists a $\pi \in S_{n-k}$ such that $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = \left(b_{\pi(1)}, b_{\pi(2)}, \ldots, b_{\pi(n-k)}\right)$. Denote this π by β. Thus, β is an element of S_{n-k} and satisfies

$$(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = \left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right). \qquad (804)$$

Now, let us introduce a notation: If $(x_1, x_2, \ldots, x_u)$ and $(y_1, y_2, \ldots, y_v)$ are two lists, then we shall let $(x_1, x_2, \ldots, x_u) * (y_1, y_2, \ldots, y_v)$ denote the list $(x_1, x_2, \ldots, x_u, y_1, y_2, \ldots, y_v)$. Then,

$$\begin{aligned}
&(\gamma(1), \gamma(2), \ldots, \gamma(n)) \\
&= \underbrace{(\gamma(1), \gamma(2), \ldots, \gamma(k))}_{\substack{=\left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right) \\ \text{(by (803))}}} * \underbrace{(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n))}_{\substack{=\left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) \\ \text{(by (804))}}} \\
&= \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}\right) * \left(b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) \\
&= \left(a_{\alpha(1)}, a_{\alpha(2)}, \ldots, a_{\alpha(k)}, b_{\beta(1)}, b_{\beta(2)}, \ldots, b_{\beta(n-k)}\right) \\
&= \left(\sigma_{I,\alpha,\beta}(1), \sigma_{I,\alpha,\beta}(2), \ldots, \sigma_{I,\alpha,\beta}(n)\right) \qquad \text{(by (799))}.
\end{aligned}$$

In other words, $\gamma(i) = \sigma_{I,\alpha,\beta}(i)$ for every $i \in \{1,2,\ldots,n\}$. In other words, $\gamma = \sigma_{I,\alpha,\beta}$ (since γ and $\sigma_{I,\alpha,\beta}$ are two maps $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$). But the definition of μ yields $\mu(\alpha,\beta) = \sigma_{I,\alpha,\beta}$. Comparing this with $\gamma = \sigma_{I,\alpha,\beta}$, we obtain

$$\gamma = \mu \underbrace{(\alpha, \beta)}_{\in S_k \times S_{n-k}} \in \mu(S_k \times S_{n-k}).$$

---

is well-defined.

We must show that $\gamma(k+j) \in \{1,2,\ldots,n\} \setminus I$.

Indeed, assume the contrary. Thus, we don't have $\gamma(k+j) \in \{1,2,\ldots,n\} \setminus I$. In other words, we have $\gamma(k+j) \notin \{1,2,\ldots,n\} \setminus I$. Combining $\gamma(k+j) \in \{1,2,\ldots,n\}$ with $\gamma(k+j) \notin \{1,2,\ldots,n\} \setminus I$, we obtain

$$\gamma(k+j) \in \{1,2,\ldots,n\} \setminus (\{1,2,\ldots,n\} \setminus I) \subseteq I = \gamma(\{1,2,\ldots,k\})$$

(since $\gamma(\{1,2,\ldots,k\}) = I$). In other words, there exists some $v \in \{1,2,\ldots,k\}$ such that $\gamma(k+j) = \gamma(v)$. Consider this v. Since $\gamma(k+j) = \gamma(v)$, we have $k+j = v$ (because γ is injective). Thus, $k+j = v \in \{1,2,\ldots,k\}$, so that $k+j \leq k$. Therefore, $j \leq k-k = 0$. This contradicts $j \in \{1,2,\ldots,n-k\}$. This contradiction shows that our assumption was wrong. Hence, $\gamma(k+j) \in \{1,2,\ldots,n\} \setminus I$ is proven. Qed.

Now, forget that we fixed γ. We thus have shown that every $\gamma \in \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\}$ satisfies $\gamma \in \mu(S_k \times S_{n-k})$. In other words,

$$\{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\} \subseteq \mu(S_k \times S_{n-k}).$$

In other words, the map μ is surjective.

So the map μ is both injective and surjective. In other words, μ is bijective. In other words, μ is a bijection. In other words, the map

$$S_k \times S_{n-k} \to \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\},$$
$$(\alpha, \beta) \mapsto \sigma_{I,\alpha,\beta}$$

is a bijection[405]. This solves Exercise 5.14 **(c)**. □

7.54. Solution to Exercise 5.15

We prepare for solving Exercise 5.15 by proving a variety of lemmas of varying triviality (all of which are simple). Our first lemma is a simple property of the permutations $t_{i,j}$ defined in Definition 5.36:

Lemma 7.93. Let X be a set. Let i and j be two distinct elements of X.

(a) We have $t_{i,j}(i) = j$.

(b) We have $t_{i,j}(j) = i$.

(c) We have $t_{i,j}(k) = k$ for each $k \in X \setminus \{i, j\}$.

(d) We have $t_{i,j} \circ t_{i,j} = \mathrm{id}$.

Proof of Lemma 7.93. Lemma 7.93 follows from the definition of $t_{i,j}$. □

Lemma 7.94. Let X and Y be two sets. Let i and j be two distinct elements of X. Let $f : X \to Y$ be a bijection.

(a) Then, $f(i)$ and $f(j)$ are two distinct elements of Y, so that the transposition $t_{f(i),f(j)}$ of Y is well-defined.

(b) This transposition satisfies $t_{f(i),f(j)} = f \circ t_{i,j} \circ f^{-1}$.

Proof of Lemma 7.94. Part **(a)** follows from the injectivity of f. Part **(b)** is verified by directly checking that $t_{f(i),f(j)}(y) = \left(f \circ t_{i,j} \circ f^{-1}\right)(y)$ for each $y \in Y$. □

[405] since μ is the map

$$S_k \times S_{n-k} \to \{\tau \in S_n \mid \tau(\{1,2,\ldots,k\}) = I\},$$
$$(\alpha, \beta) \mapsto \sigma_{I,\alpha,\beta}$$

Corollary 7.95. Let X be a finite set. Let i and j be two distinct elements of X. Then, $(-1)^{t_{i,j}} = -1$.

Proof of Corollary 7.95. Define an $n \in \mathbb{N}$ by $n = |X|$. (This is well-defined, since X is a finite set.) Then, $|X| = n = |\{1, 2, \ldots, n\}|$ (since $|\{1, 2, \ldots, n\}| = n$). Thus, there exists a bijection $\phi : X \to \{1, 2, \ldots, n\}$. Consider such a ϕ.

Let σ be the permutation $t_{i,j}$ of X. Thus, $\sigma = t_{i,j}$. Consider the number $(-1)^{\sigma}_{\phi}$ defined as in Exercise 5.12. Then, the definition of $(-1)^{\sigma}$ (in Exercise 5.12) yields $(-1)^{\sigma} = (-1)^{\sigma}_{\phi} = (-1)^{\phi \circ \sigma \circ \phi^{-1}}$ (by the definition of $(-1)^{\sigma}_{\phi}$).

Lemma 7.94 **(a)** (applied to $Y = \{1, 2, \ldots, n\}$ and $f = \phi$) shows that $\phi(i)$ and $\phi(j)$ are two distinct elements of $\{1, 2, \ldots, n\}$, so that the transposition $t_{\phi(i),\phi(j)}$ of $\{1, 2, \ldots, n\}$ is well-defined. Lemma 7.94 **(b)** (applied to $Y = \{1, 2, \ldots, n\}$ and $f = \phi$) shows that this transposition satisfies $t_{\phi(i),\phi(j)} = \phi \circ t_{i,j} \circ \phi^{-1}$. But Exercise 5.10 **(b)** (applied to $\phi(i)$ and $\phi(j)$ instead of i and j) shows that $(-1)^{t_{\phi(i),\phi(j)}} = -1$. In view of $t_{\phi(i),\phi(j)} = \phi \circ t_{i,j} \circ \phi^{-1}$, this rewrites as $(-1)^{\phi \circ t_{i,j} \circ \phi^{-1}} = -1$.

Now,

$$\begin{aligned} (-1)^{\sigma} &= (-1)^{\sigma}_{\phi} = (-1)^{\phi \circ \sigma \circ \phi^{-1}} = (-1)^{\phi \circ t_{i,j} \circ \phi^{-1}} \qquad \left(\text{since } \sigma = t_{i,j}\right) \\ &= -1. \end{aligned}$$

In view of $\sigma = t_{i,j}$, this rewrites as $(-1)^{t_{i,j}} = -1$. This proves Corollary 7.95. □

Lemma 7.96. Let X and Y be two sets. Let $f : X \to Y$ be a bijection. Let $p \in \mathbb{N}$. Let $\alpha_1, \alpha_2, \ldots, \alpha_p$ be p maps $Y \to Y$. For each $i \in \{1, 2, \ldots, p\}$, define a map $\beta_i : X \to X$ by $\beta_i = f^{-1} \circ \alpha_i \circ f$. Then,

$$\beta_1 \circ \beta_2 \circ \cdots \circ \beta_p = f^{-1} \circ \left(\alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_p\right) \circ f.$$

Proof of Lemma 7.96. We claim that

$$\beta_1 \circ \beta_2 \circ \cdots \circ \beta_m = f^{-1} \circ (\alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_m) \circ f \tag{805}$$

for each $m \in \{0, 1, \ldots, p\}$.

Indeed, (805) can be proven by a straightforward induction on m (using the definition of β_i).

Now, applying (805) to $m = p$, we find $\beta_1 \circ \beta_2 \circ \cdots \circ \beta_p = f^{-1} \circ \left(\alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_p\right) \circ f$. This proves Lemma 7.96. □

The following fact is clear from the definitions:

Proposition 7.97. Let $n \in \mathbb{N}$. Let $i \in \{1, 2, \ldots, n-1\}$. Then, $s_i = t_{i,i+1}$.

Proposition 7.98. Let X be a finite set. Let τ be any permutation of X. Then, τ can be written as a composition of finitely many transpositions of X.

Proof of Proposition 7.98. Define an $n \in \mathbb{N}$ by $n = |X|$. (This is well-defined, since X is a finite set.) Then, there exists a bijection $f : X \to \{1, 2, \ldots, n\}$. Consider such a f.

Let Y be the set $\{1, 2, \ldots, n\}$. Thus, $Y = \{1, 2, \ldots, n\}$. Hence, f is a bijection $X \to Y$ (since f is a bijection $X \to \{1, 2, \ldots, n\}$).

But τ is a permutation of X. In other words, τ is a bijection $X \to X$.

The maps f, τ and f^{-1} are bijections. Hence, their composition $f \circ \tau \circ f^{-1} : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ is also a bijection. Denote this bijection $f \circ \tau \circ f^{-1}$ by σ. Thus, $\sigma = f \circ \tau \circ f^{-1}$.

We know that σ is a bijection $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. In other words, σ is a permutation of the set $\{1, 2, \ldots, n\}$. In other words, $\sigma \in S_n$. Hence, Exercise 5.1 **(b)** shows that σ can be written as a composition of several permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). In other words, there exist some $p \in \mathbb{N}$ and some elements $k_1, k_2, \ldots, k_p$ of $\{1, 2, \ldots, n-1\}$ such that $\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p}$. Consider this p and these $k_1, k_2, \ldots, k_p$.

For each $i \in \{1, 2, \ldots, p\}$, the map s_{k_i} is a bijection $Y \to Y$ [406]. Hence, for each $i \in \{1, 2, \ldots, p\}$, we can define a map $\beta_i : X \to X$ by $\beta_i = f^{-1} \circ s_{k_i} \circ f$ (because f is a bijection $X \to Y$). Consider these β_i. Lemma 7.96 (applied to $\alpha_i = s_{k_i}$) yields that

$$\beta_1 \circ \beta_2 \circ \cdots \circ \beta_p = f^{-1} \circ \underbrace{\left(s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_p} \right)}_{= \sigma = f \circ \tau \circ f^{-1}} \circ f = \underbrace{f^{-1} \circ f}_{= \mathrm{id}} \circ \tau \circ \underbrace{f^{-1} \circ f}_{= \mathrm{id}} = \tau.$$

But for each $i \in \{1, 2, \ldots, p\}$, the map β_i is a transposition of X [407]. In other words, $\beta_1, \beta_2, \ldots, \beta_p$ are transpositions of X. Thus, the permutation $\beta_1 \circ \beta_2 \circ \cdots \circ \beta_p$

[406] *Proof.* Let $i \in \{1, 2, \ldots, p\}$. Then, $s_{k_i} \in S_n$. In other words, s_{k_i} is a permutation of the set $\{1, 2, \ldots, n\}$. In other words, s_{k_i} is a permutation of the set Y (since $Y = \{1, 2, \ldots, n\}$). In other words, s_{k_i} is a bijection $Y \to Y$. Qed.

[407] *Proof.* Let $i \in \{1, 2, \ldots, p\}$. Thus, the definition of β_i yields $\beta_i = f^{-1} \circ s_{k_i} \circ f$.

We have $k_i \in \{1, 2, \ldots, n-1\}$. Hence, Proposition 7.97 (applied to k_i instead of i) shows that $s_{k_i} = t_{k_i, k_i+1}$.

But $k_i \in \{1, 2, \ldots, n-1\}$. Hence, k_i and $k_i + 1$ are two elements of $\{1, 2, \ldots, n\} = Y$. Thus, $f^{-1}(k_i)$ and $f^{-1}(k_i + 1)$ are two elements of X (since $f : X \to Y$ is a bijection). These two elements $f^{-1}(k_i)$ and $f^{-1}(k_i + 1)$ are distinct (because if we had $f^{-1}(k_i) = f^{-1}(k_i + 1)$, then we would have $k_i = k_i + 1$, which is absurd). Thus, the map $t_{f^{-1}(k_i), f^{-1}(k_i+1)}$ is a transposition of X (by the definition of a "transposition of X").

But Lemma 7.94 **(b)** (applied to $f^{-1}(k_i)$ and $f^{-1}(k_i + 1)$ instead of i and j) shows that the transposition $t_{f(f^{-1}(k_i)), f(f^{-1}(k_i+1))}$ satisfies $t_{f(f^{-1}(k_i)), f(f^{-1}(k_i+1))} = f \circ t_{f^{-1}(k_i), f^{-1}(k_i+1)} \circ f^{-1}$. Thus,

$$\begin{aligned} f \circ t_{f^{-1}(k_i), f^{-1}(k_i+1)} \circ f^{-1} &= t_{f(f^{-1}(k_i)), f(f^{-1}(k_i+1))} \\ &= t_{k_i, k_i+1} \qquad \left(\text{since } f\left(f^{-1}(k_i)\right) = k_i \text{ and } f\left(f^{-1}(k_i+1)\right) = k_i + 1 \right) \\ &= s_{k_i} \qquad \left(\text{since } s_{k_i} = t_{k_i, k_i+1} \right). \end{aligned}$$

is a composition of finitely many transpositions of X. In view of $\beta_1 \circ \beta_2 \circ \cdots \circ \beta_p = \tau$, this rewrites as follows: The permutation τ is a composition of finitely many transpositions of X. This proves Proposition 7.98. □

Proposition 7.99. Let X be a finite set. Let $k \in \mathbb{N}$. Let $\sigma_1, \sigma_2, \ldots, \sigma_k$ be k permutations of X. Then,

$$(-1)^{\sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_k} = (-1)^{\sigma_1} \cdot (-1)^{\sigma_2} \cdot \cdots \cdot (-1)^{\sigma_k}.$$

Proof of Proposition 7.99. This is analogous to the proof of Proposition 5.28, using Exercise 5.12 **(c)**. □

Corollary 7.100. Let X be a finite set. Let $k \in \mathbb{N}$. Let σ be a permutation of X. Assume that σ can be written as a composition of k transpositions of X. Then, $(-1)^\sigma = (-1)^k$.

Proof of Corollary 7.100. We know that σ can be written as a composition of k transpositions of X. In other words, there exist k transpositions $u_1, u_2, \ldots, u_k$ of X such that $\sigma = u_1 \circ u_2 \circ \cdots \circ u_k$. Consider these $u_1, u_2, \ldots, u_k$.

Hence, $u_1, u_2, \ldots, u_k$ are k permutations of X.

Thus, Proposition 7.99 (applied to $\sigma_i = u_i$) yields

$$(-1)^{u_1 \circ u_2 \circ \cdots \circ u_k} = (-1)^{u_1} \cdot (-1)^{u_2} \cdot \cdots \cdot (-1)^{u_k} = \prod_{p=1}^{k} (-1)^{u_p}.$$

Each $p \in \{1, 2, \ldots, k\}$ satisfies $(-1)^{u_p} = -1$ [408]. Thus, $\prod_{p=1}^{k} \underbrace{(-1)^{u_p}}_{=-1} = \prod_{p=1}^{k} (-1) = (-1)^k$.

From $\sigma = u_1 \circ u_2 \circ \cdots \circ u_k$, we obtain $(-1)^\sigma = (-1)^{u_1 \circ u_2 \circ \cdots \circ u_k} = \prod_{p=1}^{k} (-1)^{u_p} = (-1)^k$. This proves Corollary 7.100. □

Hence,

$$f^{-1} \circ \underbrace{f \circ t_{f^{-1}(k_i), f^{-1}(k_i+1)} \circ f^{-1}}_{=s_{k_i}} \circ f = f^{-1} \circ s_{k_i} \circ f = \beta_i \qquad \left(\text{since } \beta_i = f^{-1} \circ s_{k_i} \circ f\right).$$

Thus, $\beta_i = \underbrace{f^{-1} \circ f}_{=\mathrm{id}_X} \circ t_{f^{-1}(k_i), f^{-1}(k_i+1)} \circ \underbrace{f^{-1} \circ f}_{=\mathrm{id}_X} = t_{f^{-1}(k_i), f^{-1}(k_i+1)}$.

But we know that the map $t_{f^{-1}(k_i), f^{-1}(k_i+1)}$ is a transposition of X. In view of $\beta_i = t_{f^{-1}(k_i), f^{-1}(k_i+1)}$, this rewrites as follows: The map β_i is a transposition of X. Qed.

[408] *Proof.* Let $p \in \{1, 2, \ldots, k\}$. Then, u_p is a transposition of X (since $u_1, u_2, \ldots, u_k$ are k transpositions of X). In other words, u_p has the form $u_p = t_{i,j}$ where i and j are two distinct elements of X (by the definition of a "transposition of X"). Consider these i and j. From $u_p = t_{i,j}$, we obtain $(-1)^{u_p} = (-1)^{t_{i,j}} = -1$ (by Corollary 7.95). Qed.

Solution to Exercise 5.15. Exercise 5.15 **(a)** is precisely Corollary 7.95. Exercise 5.15 **(b)** is precisely Proposition 7.98. Exercise 5.15 **(c)** is precisely Corollary 7.100. Thus, Exercise 5.15 is solved. □

7.55. Solution to Exercise 5.16

Throughout this section, we shall use Definition 5.29. We shall also use the notation $[n]$ for the set $\{1, 2, \ldots, n\}$ whenever $n \in \mathbb{N}$.

7.55.1. The "moving lemmas"

Before we start solving Exercise 5.16, let us state three lemmas that help us understand how a composition of several maps acts on an element. All three of these lemmas are intuitively obvious; the length of their formal proofs is mainly due to the amount of bookkeeping they require.

Lemma 7.101. Let X be a set. Let $m \in \mathbb{N}$. Let $f_1, f_2, \ldots, f_m$ be m maps from X to X. Let $x \in X$.

Assume that

$$f_j(x) = x \qquad \text{for each } j \in \{1, 2, \ldots, m\}. \tag{806}$$

Then, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = x$.

Proof of Lemma 7.101. Here is an informal proof: Imagine the element x undergoing the maps $f_1, f_2, \ldots, f_m$ in this order; the result is, of course, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x)$. But let us look closer at the step-by-step procedure. The element is initially x. Then, the maps $f_1, f_2, \ldots, f_m$ are being applied to it in this order. The element never changes in the process (because of (806)). Thus, the final result is still x. This shows that $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = x$.

If you wish, you can easily turn this argument into a rigorous proof: We claim that each $k \in \{0, 1, \ldots, m\}$ satisfies

$$(f_k \circ f_{k-1} \circ \cdots \circ f_1)(x) = x. \tag{807}$$

The proof of (807) is a straightforward induction on k (where (806) is used in the induction step). After (807) is proven, we can apply (807) to $k = m$, and conclude that $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = x$. Thus, Lemma 7.101 is proven. □

Lemma 7.102. Let X be a set. Let $m \in \mathbb{N}$. Let $f_1, f_2, \ldots, f_m$ be m maps from X to X. Let x and y be two elements of X.

Let $i \in \{1, 2, \ldots, m\}$. Assume that $f_i(x) = y$. Assume further that

$$f_j(x) = x \qquad \text{for each } j \in \{1, 2, \ldots, m\} \text{ satisfying } j < i. \tag{808}$$

Assume also that

$$f_j(y) = y \qquad \text{for each } j \in \{1, 2, \ldots, m\} \text{ satisfying } j > i. \tag{809}$$

Then, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = y$.

Proof of Lemma 7.102. Lemma 7.102 is a more complicated variant of Lemma 7.101. Again, there is a quick informal proof: Imagine the element x undergoing the maps $f_1, f_2, \ldots, f_m$ in this order; the result is, of course, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x)$. But let us look closer at the step-by-step procedure. The element is initially x. Then, the maps $f_1, f_2, \ldots, f_m$ are being applied to it in this order. Up until the map f_i is applied, the element does not change (because of (808)). Then, the map f_i is applied, and the element becomes y (since $f_i(x) = y$). From then on, the maps $f_{i+1}, f_{i+2}, \ldots, f_m$ again leave the element unchanged (due to (809)). Thus, the final result is y. This shows that $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = y$.

Again, it is not hard to formalize this argument into a rigorous proof: For each $p \in \mathbb{Z}$, we define an element $z_p \in X$ by

$$z_p = \begin{cases} x, & \text{if } p < i; \\ y, & \text{if } p \geq i \end{cases}.$$

These elements have the following property:

$$f_j(z_{j-1}) = z_j \qquad \text{for each } j \in \{1, 2, \ldots, m\}. \tag{810}$$

(Indeed, this is proven mechanically by distinguishing the cases $j < i$, $j = i$ and $j > i$. In the case when $j < i$, the equality (810) follows from (808). In the case when $j = i$, the equality (810) follows from $f_i(x) = y$. In the case when $j > i$, the equality (810) follows from (809).)

Notice that $i \in \{1, 2, \ldots, m\}$, so that $0 < i$ and $m \geq i$. The definition of z_0 yields $z_0 = x$ (since $0 < i$), whereas the definition of z_m yields $z_m = y$ (since $m \geq i$).

Now, we claim that each $k \in \{0, 1, \ldots, m\}$ satisfies

$$(f_k \circ f_{k-1} \circ \cdots \circ f_1)(x) = z_k. \tag{811}$$

The proof of (811) is a straightforward induction on k (where $z_0 = x$ is used in the induction base, and (810) is used in the induction step). After (811) is proven, we can apply (811) to $k = m$, and conclude that $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = z_m = y$. Thus, Lemma 7.102 is proven. □

The next lemma is a generalization of Lemma 7.101 (and, with a bit more work, of Lemma 7.102):

Lemma 7.103. Let X be a set. Let $m \in \mathbb{N}$. Let $f_1, f_2, \ldots, f_m$ be m maps from X to X. Let $x_0, x_1, \ldots, x_m$ be $m + 1$ elements of X.

Assume that

$$f_j(x_{j-1}) = x_j \qquad \text{for each } j \in \{1, 2, \ldots, m\}. \tag{812}$$

Then, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x_0) = x_m$.

Proof of Lemma 7.103. Here is an informal proof: From (812), we obtain the m equalities

$$f_1(x_0) = x_1, \qquad f_2(x_1) = x_2, \qquad \ldots, \qquad f_m(x_{m-1}) = x_m.$$

Now, imagine the maps $f_1, f_2, \ldots, f_m$ being applied (in this order) to the element x_0; the result is, of course, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x_0)$. But let us look closer at this step-by-step procedure. We start with the element x_0. Then, we apply the map f_1 to it, and it becomes x_1 (since $f_1(x_0) = x_1$). Next, we apply the map f_2 to it (i.e., to x_1, not to x_0), and it becomes x_2 (since $f_2(x_1) = x_2$). Then, we apply the map f_3 to it, and it becomes x_3 (for similar reasons). We then continue with $f_4, f_5, \ldots, f_m$, obtaining the elements $x_4, x_5, \ldots, x_m$ in the process (for the same reasons); hence, the final result is x_m. Thus, we know that the final result is simultaneously $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x_0)$ and x_m. Hence, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x_0) = x_m$.

If you wish, you can easily turn this argument into a rigorous proof: We claim that each $k \in \{0, 1, \ldots, m\}$ satisfies

$$(f_k \circ f_{k-1} \circ \cdots \circ f_1)(x_0) = x_k. \tag{813}$$

The proof of (813) is a straightforward induction on k (where (812) is used in the induction step). After (813) is proven, we can apply (813) to $k = m$, and conclude that $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x_0) = x_m$. Thus, Lemma 7.103 is proven. $\square$

7.55.2. Solving Exercise 5.16

Solution to Exercise 5.16. We have $k \in \{1, 2, \ldots, n\}$, thus $k \geq 1$, hence $k - 1 \geq 0$. Therefore, $k - 1 \in \mathbb{N}$.

We have defined $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ to be the permutation in S_n which sends $i_1, i_2, \ldots, i_k$ to $i_2, i_3, \ldots, i_k, i_1$, respectively, while leaving all other elements of $[n]$ fixed. Therefore:

- The permutation $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ sends $i_1, i_2, \ldots, i_k$ to $i_2, i_3, \ldots, i_k, i_1$, respectively. In other words,

$$\operatorname{cyc}_{i_1, i_2, \ldots, i_k}(i_p) = i_{p+1} \qquad \text{for every } p \in \{1, 2, \ldots, k\}, \tag{814}$$

  where i_{k+1} means i_1.

- The permutation $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ leaves all other elements of $[n]$ fixed (where "other" means "other than $i_1, i_2, \ldots, i_k$"). In other words,

$$\operatorname{cyc}_{i_1, i_2, \ldots, i_k}(q) = q \qquad \text{for every } q \in [n] \setminus \{i_1, i_2, \ldots, i_k\}. \tag{815}$$

In the following, let i_{k+1} mean i_1. Thus, $i_{k+1} = i_1$.

Define $\alpha \in S_n$ by $\alpha = \operatorname{cyc}_{i_1,i_2,\ldots,i_k}$. Define $\beta \in S_n$ by $\beta = t_{i_1,i_2} \circ t_{i_2,i_3} \circ \cdots \circ t_{i_{k-1},i_k}$.

For each $j \in \{1,2,\ldots,k-1\}$, we define a transposition g_j in S_n by $g_j = t_{i_j,i_{j+1}}$. This transposition g_j belongs to S_n, and thus is a bijective map from $[n]$ to $[n]$.

For each $j \in \{1,2,\ldots,k-1\}$, we have $k-j \in \{1,2,\ldots,k-1\}$, and therefore g_{k-j} is a well-defined bijective map from $[n]$ to $[n]$. Hence, for each $j \in \{1,2,\ldots,k-1\}$, we can define a bijective map f_j from $[n]$ to $[n]$ by $f_j = g_{k-j}$. Consider these bijective maps f_j for all $j \in \{1,2,\ldots,k-1\}$. We thus have defined $k-1$ bijective maps $f_1, f_2, \ldots, f_{k-1}$ from $[n]$ to $[n]$. Note that

$$\begin{aligned} f_{k-1} \circ f_{(k-1)-1} \circ \cdots \circ f_1 = f_{k-1} \circ f_{k-2} \circ \cdots \circ f_1 &= \underbrace{g_{k-(k-1)}}_{=g_1} \circ \underbrace{g_{k-(k-2)}}_{=g_2} \circ \cdots \circ g_{k-1} \\ &\qquad \left(\text{since } f_j = g_{k-j} \text{ for each } j \in \{1,2,\ldots,k-1\}\right) \\ &= g_1 \circ g_2 \circ \cdots \circ g_{k-1} = t_{i_1,i_2} \circ t_{i_2,i_3} \circ \cdots \circ t_{i_{k-1},i_k} \\ &\qquad \left(\text{since } g_j = t_{i_j,i_{j+1}} \text{ for each } j \in \{1,2,\ldots,k-1\}\right) \\ &= \beta \end{aligned} \tag{816}$$

(since $\beta = t_{i_1,i_2} \circ t_{i_2,i_3} \circ \cdots \circ t_{i_{k-1},i_k}$).

It is easy to see that each $j \in \{1,2,\ldots,k-1\}$ satisfies

$$f_j\left(i_{k-j}\right) = i_{k-j+1} \qquad \text{and} \tag{817}$$

$$f_j\left(i_{k-j+1}\right) = i_{k-j} \qquad \text{and} \tag{818}$$

$$\left(f_j(q) = q \qquad \text{for each } q \in [n] \setminus \left\{i_{k-j}, i_{k-j+1}\right\}\right). \tag{819}$$

[*Proof of (817), (818) and (819):* Let $j \in \{1,2,\ldots,k-1\}$. We want to prove the three claims (817), (818) and (819).

We have $f_j = g_{k-j}$ (by the definition of f_j). Moreover, $g_{k-j} = t_{i_{k-j},i_{(k-j)+1}}$ (by the definition of g_{k-j}). Thus, $f_j = g_{k-j} = t_{i_{k-j},i_{(k-j)+1}} = t_{i_{k-j},i_{k-j+1}}$ (since $(k-j)+1 = k-j+1$).

The definition of this transposition $t_{i_{k-j},i_{k-j+1}}$ shows that $t_{i_{k-j},i_{k-j+1}}$ is the permutation in S_n which switches i_{k-j} with i_{k-j+1} while leaving all other elements of $\{1,2,\ldots,n\}$ unchanged. In other words, f_j is the permutation in S_n which switches i_{k-j} with i_{k-j+1} while leaving all other elements of $[n]$ unchanged (since $f_j = t_{i_{k-j},i_{k-j+1}}$ and $[n] = \{1,2,\ldots,n\}$). In other words, f_j is the permutation in S_n that satisfies $f_j\left(i_{k-j}\right) = i_{k-j+1}$ and $f_j\left(i_{k-j+1}\right) = i_{k-j}$ and

$$\left(f_j(q) = q \qquad \text{for each } q \in [n] \setminus \left\{i_{k-j}, i_{k-j+1}\right\}\right).$$

Thus, the three claims (817), (818) and (819) are proven.]

Now, we claim that each $x \in [n]$ satisfies

$$\alpha(x) = \beta(x). \tag{820}$$

[*Proof of (820):* Let $x \in [n]$. We are in one of the following three cases:

Case 1: We have $x \notin \{i_1, i_2, \ldots, i_k\}$.

Case 2: We have $x = i_k$.

Case 3: We have neither $x \notin \{i_1, i_2, \ldots, i_k\}$ nor $x = i_k$.

Let us first consider Case 1. In this case, we have $x \notin \{i_1, i_2, \ldots, i_k\}$. Combining $x \in [n]$ with $x \notin \{i_1, i_2, \ldots, i_k\}$, we obtain $x \in [n] \setminus \{i_1, i_2, \ldots, i_k\}$. Hence, (815) (applied to $q = x$) yields $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}(x) = x$. On the other hand, we have $f_j(x) = x$ for each $j \in \{1, 2, \ldots, k-1\}$ [409]. Hence, Lemma 7.101 (applied to $m = k-1$ and $X = [n]$) yields $\left(f_{k-1} \circ f_{(k-1)-1} \circ \cdots \circ f_1\right)(x) = x$. In view of (816), this rewrites as $\beta(x) = x$. Comparing this with $\underbrace{\alpha}_{=\operatorname{cyc}_{i_1,i_2,\ldots,i_k}}(x) = \operatorname{cyc}_{i_1,i_2,\ldots,i_k}(x) = x$, we obtain $\alpha(x) = \beta(x)$. Hence, (820) is proven in Case 1.

Let us next consider Case 2. In this case, we have $x = i_k$.

We have $f_j\left(i_{k-(j-1)}\right) = i_{k-j}$ for each $j \in \{1, 2, \ldots, k-1\}$ [410]. Thus, Lemma 7.103 (applied to $m = k-1$ and $X = [n]$ and $x_j = i_{k-j}$) yields $\left(f_{k-1} \circ f_{(k-1)-1} \circ \cdots \circ f_1\right)(i_{k-0}) = i_{k-(k-1)}$. In view of $k - 0 = k$ and $k - (k-1) = 1$, this rewrites as $\left(f_{k-1} \circ f_{(k-1)-1} \circ \cdots \circ f_1\right)(i_k) = i_1$. In view of (816), this rewrites as $\beta(i_k) = i_1$.

On the other hand, $k \in \{1, 2, \ldots, k\}$ (since $k \geq 1$). Hence, (814) (applied to $p = k$) yields $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}(i_k) = i_{k+1} = i_1$. Thus,

$$\underbrace{\alpha}_{=\operatorname{cyc}_{i_1,i_2,\ldots,i_k}}\left(\underbrace{x}_{=i_k}\right) = \operatorname{cyc}_{i_1,i_2,\ldots,i_k}(i_k) = i_1.$$

Comparing this with $\beta\left(\underbrace{x}_{=i_k}\right) = \beta(i_k) = i_1$, we obtain $\alpha(x) = \beta(x)$. Hence, (820) is proven in Case 2.

Let us finally consider Case 3. In this case, we have neither $x \notin \{i_1, i_2, \ldots, i_k\}$ nor $x = i_k$. Thus, we don't have $x \notin \{i_1, i_2, \ldots, i_k\}$. Hence, we have $x \in \{i_1, i_2, \ldots, i_k\}$. In other words, $x = i_p$ for some $p \in \{1, 2, \ldots, k\}$. Consider this p.

Also, we don't have $x = i_k$ (since we have neither $x \notin \{i_1, i_2, \ldots, i_k\}$ nor $x = i_k$). In other words, we have $x \neq i_k$. From $x = i_p$, we obtain $i_p = x \neq i_k$, thus $p \neq k$. Combining $p \in \{1, 2, \ldots, k\}$ with $p \neq k$, we obtain $p \in \{1, 2, \ldots, k\} \setminus \{k\} =$

---

[409] *Proof.* Let $j \in \{1, 2, \ldots, k-1\}$. Thus, $x \in [n] \setminus \underbrace{\{i_1, i_2, \ldots, i_k\}}_{\supseteq \{i_{k-j}, i_{k-j+1}\}} \subseteq [n] \setminus \left\{i_{k-j}, i_{k-j+1}\right\}$. Hence, (819) (applied to $q = x$) yields $f_j(x) = x$. Qed.

[410] *Proof.* Let $j \in \{1, 2, \ldots, k-1\}$. Then, (818) yields $f_j\left(i_{k-j+1}\right) = i_{k-j}$. In view of $k - j + 1 = k - (j-1)$, this rewrites as $f_j\left(i_{k-(j-1)}\right) = i_{k-j}$. Qed.

$\{1,2,\ldots,k-1\}$. Hence, $k-p\in\{1,2,\ldots,k-1\}$. Thus, the map f_{k-p} is well-defined.

We have $p\in\{1,2,\ldots,k-1\}$, thus $p+1\in\{2,3,\ldots,k\}\subseteq\{1,2,\ldots,k\}$. Hence, $i_{p+1}\in\{i_1,i_2,\ldots,i_k\}\subseteq[n]$ (since $i_1,i_2,\ldots,i_k$ are elements of $[n]$). Thus, we can define $y\in[n]$ by $y=i_{p+1}$. Consider this y. Thus,

$$x=i_p \qquad \text{and} \qquad y=i_{p+1}.$$

The equality (817) (applied to $j=k-p$) yields $f_{k-p}\left(i_{k-(k-p)}\right)=i_{k-(k-p)+1}$. In view of $k-(k-p)=p$, this rewrites as $f_{k-p}\left(i_p\right)=i_{p+1}$. In view of $i_p=x$ and $i_{p+1}=y$, this rewrites as

$$f_{k-p}(x)=y. \tag{821}$$

Moreover, we have

$$f_j(x)=x \qquad \text{for each } j\in\{1,2,\ldots,k-1\} \text{ satisfying } j<k-p. \tag{822}$$

[*Proof of (822):* Let $j\in\{1,2,\ldots,k-1\}$ be such that $j<k-p$.

From $j<k-p$, we obtain $k-j>p$, thus $k-j\neq p$. Hence, $i_{k-j}\neq i_p$ (since the k elements $i_1,i_2,\ldots,i_k$ are distinct). In other words, $i_p\neq i_{k-j}$. Hence, $x=i_p\neq i_{k-j}$.

We have $k-j+1>k-j>p$, thus $k-j+1\neq p$. Hence, $i_{k-j+1}\neq i_p$ (since the k elements $i_1,i_2,\ldots,i_k$ are distinct). In other words, $i_p\neq i_{k-j+1}$. Hence, $x=i_p\neq i_{k-j+1}$.

Combining $x\neq i_{k-j}$ with $x\neq i_{k-j+1}$, we obtain $x\notin\left\{i_{k-j},i_{k-j+1}\right\}$. Combining $x\in[n]$ with this, we obtain $x\in[n]\setminus\left\{i_{k-j},i_{k-j+1}\right\}$. Hence, (819) (applied to $q=x$) yields $f_j(x)=x$. This proves (822).]

Furthermore, we have

$$f_j(y)=y \qquad \text{for each } j\in\{1,2,\ldots,k-1\} \text{ satisfying } j>k-p. \tag{823}$$

[*Proof of (823):* Let $j\in\{1,2,\ldots,k-1\}$ be such that $j>k-p$.

From $j>k-p$, we obtain $k-j<p<p+1$, thus $k-j\neq p+1$. Hence, $i_{k-j}\neq i_{p+1}$ (since the k elements $i_1,i_2,\ldots,i_k$ are distinct). In other words, $i_{p+1}\neq i_{k-j}$. Hence, $y=i_{p+1}\neq i_{k-j}$.

We have $k-j<p$, thus $k-j\neq p$ and therefore $k-j+1\neq p+1$. Hence, $i_{k-j+1}\neq i_{p+1}$ (since the k elements $i_1,i_2,\ldots,i_k$ are distinct). In other words, $i_{p+1}\neq i_{k-j+1}$. Hence, $y=i_{p+1}\neq i_{k-j+1}$.

Combining $y\neq i_{k-j}$ with $y\neq i_{k-j+1}$, we obtain $y\notin\left\{i_{k-j},i_{k-j+1}\right\}$. Combining $y\in[n]$ with this, we obtain $y\in[n]\setminus\left\{i_{k-j},i_{k-j+1}\right\}$. Hence, (819) (applied to $q=y$) yields $f_j(y)=y$. This proves (823).]

Now, we have proven the equalities (821), (822) and (823). Hence, Lemma 7.102 (applied to $k-1$, $[n]$ and $k-p$ instead of m, X and i) yields

$$\left(f_{k-1}\circ f_{(k-1)-1}\circ\cdots\circ f_1\right)(x)=y.$$

In view of (816), this rewrites as $\beta(x) = y$. Comparing this with

$$\underbrace{\alpha}_{=\operatorname{cyc}_{i_1,i_2,\ldots,i_k}}\left(\underbrace{x}_{=i_p}\right) = \operatorname{cyc}_{i_1,i_2,\ldots,i_k}(i_p) = i_{p+1} \qquad \text{(by (814))}$$
$$= y,$$

we obtain $\alpha(x) = \beta(x)$. Hence, (820) is proven in Case 3.

We have now proven (820) in each of the three Cases 1, 2 and 3. Since these three Cases cover all possibilities, we thus conclude that (820) always holds.]

But from (820), we immediately obtain $\alpha = \beta$ (since both α and β are maps from $[n]$ to $[n]$). In view of $\alpha = \operatorname{cyc}_{i_1,i_2,\ldots,i_k}$ and $\beta = t_{i_1,i_2} \circ t_{i_2,i_3} \circ \cdots \circ t_{i_{k-1},i_k}$, this rewrites as

$$\operatorname{cyc}_{i_1,i_2,\ldots,i_k} = t_{i_1,i_2} \circ t_{i_2,i_3} \circ \cdots \circ t_{i_{k-1},i_k}.$$

This solves Exercise 5.16. □

7.55.3. A particular case

In this subsection, we shall derive a particular case of Exercise 5.16 that will come useful later. We make a definition:

Definition 7.104. Let $n \in \mathbb{N}$. Let u and v be two elements of $[n]$ such that $u \leq v$. Then, we define a permutation $c_{u,v} \in S_n$ by

$$c_{u,v} = \operatorname{cyc}_{v,v-1,v-2,\ldots,u}.$$

(This is well-defined by Corollary 7.105 **(a)** below.)

Corollary 7.105. Let $n \in \mathbb{N}$. Let u and v be two elements of $[n]$ such that $u \leq v$. Then:

(a) The permutation $c_{u,v}$ is well-defined.

(b) We have $c_{u,v} = s_{v-1} \circ s_{v-2} \circ \cdots \circ s_u$.

Before we prove this corollary, let us establish an almost trivial lemma:

Lemma 7.106. Let $n \in \mathbb{N}$. Let $i \in \{1, 2, \ldots, n-1\}$. Then, $t_{i+1,i} = s_i$.

Proof of Lemma 7.106. Recall that s_i is the permutation in S_n that switches i with $i+1$ but leaves all other numbers unchanged (by the definition of s_i).

On the other hand, $i \in \{1, 2, \ldots, n-1\}$. Hence, $i+1$ and i are two distinct elements of $\{1, 2, \ldots, n\}$. Thus, $t_{i+1,i}$ is the permutation in S_n which switches $i+1$ with i while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged (by the definition of $t_{i+1,i}$). In other words, $t_{i+1,i}$ is the permutation in S_n which switches i with $i+1$ while leaving all other elements of $\{1, 2, \ldots, n\}$ unchanged (because switching $i+1$

with i is tantamount to switching i with $i+1$). In other words, $t_{i+1,i}$ is the permutation in S_n that switches i with $i+1$ but leaves all other numbers unchanged.

Now, we have shown that both $t_{i+1,i}$ and s_i are the permutation in S_n that switches i with $i+1$ but leaves all other numbers unchanged. Hence, $t_{i+1,i}$ and s_i are the same permutation. In other words, $t_{i+1,i} = s_i$. This proves Lemma 7.106. □

Proof of Corollary 7.105. We have $u \in [n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$), thus $u \geq 1$. Hence, $1 \leq u$. Also, $v \in [n] = \{1, 2, \ldots, n\}$, thus $v \leq n$. Hence, $1 \leq u \leq v \leq n$.

Also, $\underbrace{v}_{\leq n} - \underbrace{u}_{\geq 1} \leq n-1$. Combining this with $v - \underbrace{u}_{\leq v} \geq v - v = 0$, we obtain $v - u \in \{0, 1, \ldots, n-1\}$, so that $v - u + 1 \in \{1, 2, \ldots, n\}$.

Now, $v, v-1, v-2, \ldots, u$ are $v - u + 1$ distinct elements of the set $\{1, 2, \ldots, n\}$ (since $1 \leq u \leq v \leq n$). Hence, the permutation $\operatorname{cyc}_{v,v-1,v-2,\ldots,u}$ is well-defined. In other words, the permutation $c_{u,v}$ is well-defined (since $c_{u,v}$ was defined by $c_{u,v} = \operatorname{cyc}_{v,v-1,v-2,\ldots,u}$). This proves Corollary 7.105 **(a)**.

(b) Each $i \in \{v-1, v-2, \ldots, u\}$ satisfies $t_{i+1,i} = s_i$ (by Lemma 7.106) and thus $s_i = t_{i+1,i}$. In other words,

$$\begin{aligned}(s_{v-1}, s_{v-2}, \ldots, s_u) &= \left(t_{(v-1)+1,v-1}, t_{(v-2)+1,v-2}, \ldots, t_{u+1,u}\right) \\ &= \left(t_{v,v-1}, t_{v-1,v-2}, \ldots, t_{u+1,u}\right).\end{aligned}$$

Hence,

$$s_{v-1} \circ s_{v-2} \circ \cdots \circ s_u = t_{v,v-1} \circ t_{v-1,v-2} \circ \cdots \circ t_{u+1,u}. \tag{824}$$

But Exercise 5.16 (applied to $k = v - u + 1$ and $(i_1, i_2, \ldots, i_k) = (v, v-1, v-2, \ldots, u)$) yields

$$\operatorname{cyc}_{v,v-1,v-2,\ldots,u} = t_{v,v-1} \circ t_{v-1,v-2} \circ \cdots \circ t_{u+1,u} = s_{v-1} \circ s_{v-2} \circ \cdots \circ s_u$$

(by (824)). Now, the definition of $c_{u,v}$ yields $c_{u,v} = \operatorname{cyc}_{v,v-1,v-2,\ldots,u} = s_{v-1} \circ s_{v-2} \circ \cdots \circ s_u$. This proves Corollary 7.105 **(b)**. □

7.56. Solution to Exercise 5.17

Solution to Exercise 5.17. **(a)** This proof is going to be long, but most of it will be spent unraveling the notations. If you find Exercise 5.17 **(a)** obvious, don't let this proof cast doubt on your understanding.

Let $\sigma \in S_n$. Let $i_1, i_2, \ldots, i_k$ be k distinct elements of $[n]$.

The map σ is a permutation (since $\sigma \in S_n$), and therefore bijective. Hence, in particular, σ is injective.

For every $p \in \{1, 2, \ldots, k\}$, let j_p be the element $\sigma(i_p) \in [n]$. Then, $(j_1, j_2, \ldots, j_k) = (\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k))$. Furthermore, $j_1, j_2, \ldots, j_k$ are k distinct elements of $[n]$ [411]. Therefore, $\operatorname{cyc}_{j_1, j_2, \ldots, j_k}$ is a well-defined permutation in S_n.

We have defined $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ to be the permutation in S_n which sends $i_1, i_2, \ldots, i_k$ to $i_2, i_3, \ldots, i_k, i_1$, respectively, while leaving all other elements of $[n]$ fixed. Therefore:

- The permutation $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ sends $i_1, i_2, \ldots, i_k$ to $i_2, i_3, \ldots, i_k, i_1$, respectively. In other words,

$$\operatorname{cyc}_{i_1, i_2, \ldots, i_k}(i_p) = i_{p+1} \qquad \text{for every } p \in \{1, 2, \ldots, k\}, \tag{825}$$

  where i_{k+1} means i_1.

- The permutation $\operatorname{cyc}_{i_1, i_2, \ldots, i_k}$ leaves all other elements of $[n]$ fixed (where "other" means "other than $i_1, i_2, \ldots, i_k$"). In other words,

$$\operatorname{cyc}_{i_1, i_2, \ldots, i_k}(q) = q \qquad \text{for every } q \in [n] \setminus \{i_1, i_2, \ldots, i_k\}. \tag{826}$$

Similarly, we can say the same about $\operatorname{cyc}_{j_1, j_2, \ldots, j_k}$:

- We have

$$\operatorname{cyc}_{j_1, j_2, \ldots, j_k}(j_p) = j_{p+1} \qquad \text{for every } p \in \{1, 2, \ldots, k\}, \tag{827}$$

  where j_{k+1} means j_1.

- We have

$$\operatorname{cyc}_{j_1, j_2, \ldots, j_k}(q) = q \qquad \text{for every } q \in [n] \setminus \{j_1, j_2, \ldots, j_k\}. \tag{828}$$

In the following, we shall use the notation i_{k+1} as a synonym for i_1, and the notation j_{k+1} as a synonym for j_1. Then,

$$j_p = \sigma(i_p) \qquad \text{for every } p \in \{1, 2, \ldots, k+1\}. \tag{829}$$

(Indeed, in the case when $p \in \{1, 2, \ldots, k\}$, this follows from the definition of j_p; but in the remaining case when $p = k+1$, it follows from $j_{k+1} = j_1 = \sigma\left(\underbrace{i_1}_{= i_{k+1}}\right) = \sigma(i_{k+1})$.)

[411] *Proof.* The map σ is injective. Hence, the elements $\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k)$ are distinct (since the elements $i_1, i_2, \ldots, i_k$ are distinct). Thus, $\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k)$ are k distinct elements of $[n]$. In other words, $j_1, j_2, \ldots, j_k$ are k distinct elements of $[n]$ (since $(j_1, j_2, \ldots, j_k) = (\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k))$).

Now, let us show that

$$\left(\sigma \circ \operatorname{cyc}_{i_1,i_2,\ldots,i_k} \circ \sigma^{-1}\right)(q) = \operatorname{cyc}_{j_1,j_2,\ldots,j_k}(q) \qquad (830)$$

for every $q \in [n]$.

[*Proof of (830):* Let $q \in [n]$. We must prove (830). We are in one of the following two cases:

Case 1: We have $q \in \{j_1, j_2, \ldots, j_k\}$.

Case 2: We have $q \notin \{j_1, j_2, \ldots, j_k\}$.

Let us first consider Case 1. In this case, we have $q \in \{j_1, j_2, \ldots, j_k\}$. Thus, $q = j_p$ for some $p \in \{1, 2, \ldots, k\}$. Consider this p. Clearly, $p + 1 \in \{2, 3, \ldots, k+1\} \subseteq \{1, 2, \ldots, k+1\}$. Hence, applying (829) to $p+1$ instead of p, we obtain $j_{p+1} = \sigma\left(i_{p+1}\right)$. But $q = j_p = \sigma\left(i_p\right)$ (by the definition of j_p) and thus $\sigma^{-1}(q) = i_p$. Hence,

$$\left(\sigma \circ \operatorname{cyc}_{i_1,i_2,\ldots,i_k} \circ \sigma^{-1}\right)(q) = \sigma\left(\operatorname{cyc}_{i_1,i_2,\ldots,i_k}\left(\underbrace{\sigma^{-1}(q)}_{=i_p}\right)\right) = \sigma\left(\underbrace{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}\left(i_p\right)}_{\substack{=i_{p+1} \\ \text{(by (825))}}}\right)$$
$$= \sigma\left(i_{p+1}\right).$$

Compared with

$$\operatorname{cyc}_{j_1,j_2,\ldots,j_k}\left(\underbrace{q}_{=j_p}\right) = \operatorname{cyc}_{j_1,j_2,\ldots,j_k}\left(j_p\right) = j_{p+1} \qquad \text{(by (827))}$$
$$= \sigma\left(i_{p+1}\right),$$

this yields $\left(\sigma \circ \operatorname{cyc}_{i_1,i_2,\ldots,i_k} \circ \sigma^{-1}\right)(q) = \operatorname{cyc}_{j_1,j_2,\ldots,j_k}(q)$. Thus, (830) is proven in Case 1.

Let us now consider Case 2. In this case, we have $q \notin \{j_1, j_2, \ldots, j_k\}$. Hence, $\sigma^{-1}(q) \notin \{i_1, i_2, \ldots, i_k\}$ [412]. Thus, $\sigma^{-1}(q) \in [n] \setminus \{i_1, i_2, \ldots, i_k\}$. Therefore, (826) (applied to $\sigma^{-1}(q)$ instead of q) yields $\operatorname{cyc}_{i_1,i_2,\ldots,i_k}\left(\sigma^{-1}(q)\right) = \sigma^{-1}(q)$. Hence,

$$\left(\sigma \circ \operatorname{cyc}_{i_1,i_2,\ldots,i_k} \circ \sigma^{-1}\right)(q) = \sigma\left(\underbrace{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}\left(\sigma^{-1}(q)\right)}_{=\sigma^{-1}(q)}\right) = \sigma\left(\sigma^{-1}(q)\right) = q. \qquad (831)$$

[412] *Proof.* Assume the contrary. Thus, $\sigma^{-1}(q) \in \{i_1, i_2, \ldots, i_k\}$. In other words, there exists a $p \in \{1, 2, \ldots, k\}$ such that $\sigma^{-1}(q) = i_p$. Consider this p. We have $\sigma^{-1}(q) = i_p$, thus $q = \sigma\left(i_p\right) = j_p$ (since j_p is defined as $\sigma\left(i_p\right)$). Thus, $q = j_p \in \{j_1, j_2, \ldots, j_k\}$; but this contradicts $q \notin \{j_1, j_2, \ldots, j_k\}$. This contradiction shows that our assumption was wrong, qed.

On the other hand, $q \in [n] \setminus \{j_1, j_2, \ldots, j_k\}$ (since $q \notin \{j_1, j_2, \ldots, j_k\}$) and thus $\operatorname{cyc}_{j_1, j_2, \ldots, j_k}(q) = q$ (by (828)). Compared with (831), this yields $\left(\sigma \circ \operatorname{cyc}_{i_1, i_2, \ldots, i_k} \circ \sigma^{-1}\right)(q) = \operatorname{cyc}_{j_1, j_2, \ldots, j_k}(q)$. Thus, (830) is proven in Case 2.

We have now proven (830) in each of the two Cases 1 and 2. Therefore, (830) always holds.]

From this, we conclude that $\sigma \circ \operatorname{cyc}_{i_1, i_2, \ldots, i_k} \circ \sigma^{-1} = \operatorname{cyc}_{j_1, j_2, \ldots, j_k}$ (because both $\sigma \circ \operatorname{cyc}_{i_1, i_2, \ldots, i_k} \circ \sigma^{-1}$ and $\operatorname{cyc}_{j_1, j_2, \ldots, j_k}$ are maps from $[n]$ to $[n]$). Hence,

$$\sigma \circ \operatorname{cyc}_{i_1, i_2, \ldots, i_k} \circ \sigma^{-1} = \operatorname{cyc}_{j_1, j_2, \ldots, j_k} = \operatorname{cyc}_{\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k)}$$

(since $(j_1, j_2, \ldots, j_k) = (\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k))$). This solves Exercise 5.17 **(a)**.

(b) This is again a straightforward argument whose complexity stems only from the number of cases that need to be considered. We shall try to reduce the amount of brainless verification using some tricks, although at the cost of making parts of the solution appear unmotivated.

Let $p \in \{0, 1, \ldots, n-k\}$. Then, the elements $p+1, p+2, \ldots, p+k$ belong to $\{1, 2, \ldots, n\} = [n]$. Hence, $\operatorname{cyc}_{p+1, p+2, \ldots, p+k}$ is well-defined. We let σ denote the permutation $\operatorname{cyc}_{p+1, p+2, \ldots, p+k}$.

The permutation $\sigma = \operatorname{cyc}_{p+1, p+2, \ldots, p+k}$ is defined to be the permutation in S_n which sends $p+1, p+2, \ldots, p+k$ to $p+2, p+3, \ldots, p+k, p+1$, respectively, while leaving all other elements of $[n]$ fixed. Therefore:

- The permutation σ sends $p+1, p+2, \ldots, p+k$ to $p+2, p+3, \ldots, p+k, p+1$, respectively. In other words, we have

$$(\sigma(p+i) = p + (i+1) \qquad \text{for every } i \in \{1, 2, \ldots, k-1\}) \tag{832}$$

and

$$\sigma(p+k) = p+1. \tag{833}$$

- The permutation σ leaves all other elements of $[n]$ fixed (where "other" means "other than $p+1, p+2, \ldots, p+k$"). In other words,

$$\sigma(q) = q \qquad \text{for every } q \in [n] \setminus \{p+1, p+2, \ldots, p+k\}. \tag{834}$$

We can now observe that

$$\sigma(q) \geq q \qquad \text{for every } q \in [n] \text{ satisfying } q \neq p+k \tag{835}$$

[413]. Furthermore,

$$\sigma(q) \leq q+1 \qquad \text{for every } q \in [n] \tag{836}$$

[413] *Proof of (835):* Let $q \in [n]$ be such that $q \neq p+k$. We must prove that $\sigma(q) \geq q$.

We are in one of the following two cases:

Case 1: We have $q \in \{p+1, p+2, \ldots, p+k\}$.

Case 2: We have $q \notin \{p+1, p+2, \ldots, p+k\}$.

[414].

Now, set
$$A = \{(p+h, p+k) \mid h \in \{1,2,\ldots,k-1\}\}.$$
In other words,
$$A = \{(p+1, p+k), (p+2, p+k), \ldots, (p+(k-1), p+k)\}.$$
Thus, the set A has $k-1$ elements (since the $k-1$ pairs $(p+1, p+k), (p+2, p+k), \ldots, (p+(k-1), p+k)$ are clearly distinct). In other words, $|A| = k-1$.

Now, let $\operatorname{Inv}\sigma$ denote the set of all inversions of σ. Recall that $\ell(\sigma)$ was defined as the number of inversions of σ. In other words, $\ell(\sigma) = |\operatorname{Inv}\sigma|$.

Let us first consider Case 1. In this case, we have $q \in \{p+1, p+2, \ldots, p+k\}$. Hence, $q = p+i$ for some $i \in \{1,2,\ldots,k\}$. Consider this i. We have $q \neq p+k$, so that $q-p \neq k$ and thus $k \neq q-p = i$ (since $q = p+i$). Therefore, $i \neq k$. Combined with $i \in \{1,2,\ldots,k\}$, this shows that $i \in \{1,2,\ldots,k\} \setminus \{k\} = \{1,2,\ldots,k-1\}$. Hence, (832) yields $\sigma(p+i) = p + \underbrace{(i+1)}_{\geq i} \geq p+i = q$.

Hence, $\sigma\left(\underbrace{q}_{=p+i}\right) = \sigma(p+i) \geq q$. We thus have proven $\sigma(q) \geq q$ in Case 1.

Let us now consider Case 2. In this case, we have $q \notin \{p+1, p+2, \ldots, p+k\}$. Hence, $q \in [n] \setminus \{p+1, p+2, \ldots, p+k\}$. Therefore, $\sigma(q) = q$ (by (834)). Thus, $\sigma(q) \geq q$ is proven in Case 2.

We have now proven $\sigma(q) \geq q$ in both Cases 1 and 2. Therefore, $\sigma(q) \geq q$ always holds. This proves (835).

[414] *Proof of (836):* Let $q \in [n]$. We must prove that $\sigma(q) \leq q+1$.

Assume the contrary (for the sake of contradiction). Thus, $\sigma(q) > q+1$. If we had $q \in [n] \setminus \{p+1, p+2, \ldots, p+k\}$, then we would have $\sigma(q) = q$ (by (834)), which would contradict $\sigma(q) > q+1 > q$. Thus, we cannot have $q \in [n] \setminus \{p+1, p+2, \ldots, p+k\}$. We therefore have
$$q \in [n] \setminus ([n] \setminus \{p+1, p+2, \ldots, p+k\}) \subseteq \{p+1, p+2, \ldots, p+k\}.$$
Hence, $q = p+i$ for some $i \in \{1,2,\ldots,k\}$. Consider this i. Clearly, $i \geq 1$ and $i \leq k$.

If we had $i = k$, then we would have
$$\sigma\left(\underbrace{q}_{=p+i}\right) = \sigma\left(p + \underbrace{i}_{=k}\right) = \sigma(p+k) = p + \underbrace{1}_{\leq i} \qquad \text{(by (833))}$$
$$\leq p+i = q,$$
which would contradict $\sigma(q) > q+1 > q$. Thus, we cannot have $i = k$. Hence, $i \in \{1,2,\ldots,k-1\}$ (since $i \geq 1$ and $i \leq k$). Therefore, (832) yields $\sigma(p+i) = p+(i+1) = \underbrace{p+i}_{=q} + 1 = q+1$. This contradicts $\sigma\left(\underbrace{p+i}_{=q}\right) = \sigma(q) > q+1$. This contradiction proves that our assumption was wrong. Hence, $\sigma(q) \leq q+1$ is proven.

But $A \subseteq \operatorname{Inv}\sigma$ [415] and $\operatorname{Inv}\sigma \subseteq A$ [416]. Combining these two relations, we obtain $A = \operatorname{Inv}\sigma$. Hence, $|A| = |\operatorname{Inv}\sigma|$. Compared with $\ell(\sigma) = |\operatorname{Inv}\sigma|$, this yields $\ell(\sigma) = |A| = k-1$. This rewrites as $\ell\left(\operatorname{cyc}_{p+1,p+2,\ldots,p+k}\right) = k-1$ (since $\sigma = \operatorname{cyc}_{p+1,p+2,\ldots,p+k}$). This solves Exercise 5.17 **(b)**.

(c) This one is tricky. Let $i_1, i_2, \ldots, i_k$ be k distinct elements of $[n]$. We extend the k-tuple $(i_1, i_2, \ldots, i_k)$ to an infinite sequence $(i_1, i_2, i_3, \ldots)$ of elements of $[n]$ by

---

[415] *Proof.* Let $c \in A$. We shall show that $c \in \operatorname{Inv}\sigma$.

We have $c \in A = \{(p+h, p+k) \mid h \in \{1,2,\ldots,k-1\}\}$. Hence, c can be written in the form $c = (p+h, p+k)$ for some $h \in \{1,2,\ldots,k-1\}$. Consider this h. From $h \in \{1,2,\ldots,k-1\}$, we obtain $1 \leq h \leq k-1$, so that $h \leq k-1 < k$. Thus, $p + \underbrace{h}_{<k} < p+k$. Moreover, $1 \leq p+h$ (since $\underbrace{p}_{\geq 0} + \underbrace{h}_{\geq 1} \geq 0+1 = 1$) and $p+k \leq n$ (since $p \leq n-k$). Thus, $1 \leq p+h < p+k \leq n$.

Applying (832) to $i = h$, we obtain $\sigma(p+h) = p + \left(\underbrace{h}_{\geq 1 > 0} + 1\right) > p + (0+1) = p+1 = \sigma(p+k)$ (by (833)).

Now, $(p+h, p+k)$ is a pair of integers satisfying $1 \leq p+h < p+k \leq n$ and $\sigma(p+h) > \sigma(p+k)$. In other words, $(p+h, p+k)$ is a pair of integers (i,j) satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. In other words, $(p+h, p+k)$ is an inversion of σ (by the definition of "inversion of σ"). In other words, $(p+h, p+k) \in \operatorname{Inv}\sigma$. Thus, $c = (p+h, p+k) \in \operatorname{Inv}\sigma$.

Now, let us forget that we fixed c. We thus have shown that $c \in \operatorname{Inv}\sigma$ for every $c \in A$. In other words, $A \subseteq \operatorname{Inv}\sigma$, qed.

[416] *Proof.* Let $c \in \operatorname{Inv}\sigma$. We shall show that $c \in A$.

We have $c \in \operatorname{Inv}\sigma$. In other words, c is an inversion of σ. In other words, c is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. Consider this (i,j).

We have $\sigma(i) > \sigma(j)$, so that $\sigma(i) \geq \sigma(j) + 1$ (since $\sigma(i)$ and $\sigma(j)$ are integers). Also, $i < j$, so that $i \leq j-1$ (since i and j are integers). In other words, $i+1 \leq j$. But (836) (applied to $q = i$) yields $\sigma(i) \leq i+1 \leq j$.

Let us first show that $j = p+k$. Indeed, let us assume the contrary (for the sake of contradiction). Thus, $j \neq p+k$. Hence, $\sigma(j) \geq j$ (by (835), applied to $q = j$). Now, $\sigma(i) > \sigma(j) \geq j$. This contradicts $\sigma(i) \leq j$. This contradiction shows that our assumption was wrong. Hence, $j = p+k$ is proven.

Now, $\sigma\left(\underbrace{j}_{=p+k}\right) = \sigma(p+k) = p+1$ (by (833)). Hence, $\sigma(i) > \sigma(j) = p+1$. Therefore, $p+1 < \sigma(i) \leq i+1$. Subtracting 1 from both sides of this inequality, we obtain $p < i$. Hence, $i > p$, so that $i \geq p+1$ (since i and p are integers). Combined with $i < j = p+k$, this yields $i \in \{p+1, p+2, \ldots, p+k-1\}$. Thus, $i - p \in \{1,2,\ldots,k-1\}$.

So we know that the element $i - p \in \{1,2,\ldots,k-1\}$ satisfies $c = \left(\underbrace{i}_{=p+(i-p)}, \underbrace{j}_{=p+k}\right) = (p+(i-p), p+k)$. Hence, there exists an $h \in \{1,2,\ldots,k-1\}$ such that $c = (p+h, p+k)$ (namely, $h = i-p$). Thus,

$$c \in \{(p+h, p+k) \mid h \in \{1,2,\ldots,k-1\}\} = A.$$

Now, let us forget that we fixed c. We thus have proven that $c \in A$ for every $c \in \operatorname{Inv}\sigma$. In other words, $\operatorname{Inv}\sigma \subseteq A$, qed.

setting

$$\left(i_u = i_{(\text{the element } u' \in \{1,2,\ldots,k\} \text{ satisfying } u' \equiv u \bmod k)} \qquad \text{for every } u \geq 1\right).$$

This sequence $(i_1, i_2, i_3, \ldots)$ is periodic with period k. In other words,

$$i_u = i_{u+k} \qquad \text{for every } u \geq 1. \tag{837}$$

From this, it is easy to obtain that

$$i_1 + i_2 + \cdots + i_k = i_2 + i_3 + \cdots + i_{k+1} = i_3 + i_4 + \cdots + i_{k+2} = \cdots$$

[417]. Thus,

$$i_{r+1} + i_{r+2} + \cdots + i_{r+k} = i_1 + i_2 + \cdots + i_k \qquad \text{for every } r \in \mathbb{N}. \tag{838}$$

Let $\sigma = \operatorname{cyc}_{i_1, i_2, \ldots, i_k}$. Let $\operatorname{Inv} \sigma$ denote the set of all inversions of σ. Then, $\ell(\sigma) = |\operatorname{Inv} \sigma|$. (This can be seen as in the solution to Exercise 5.17 **(b)**.) Moreover, the definitions of the sequence $(i_1, i_2, i_3, \ldots)$ and of σ show that

$$\sigma(i_p) = i_{p+1} \qquad \text{for every } p \geq 1. \tag{839}$$

Now, fix $r \in \{1, 2, \ldots, k-1\}$. We shall prove that

$$\text{there exists some } u \geq 1 \text{ such that } (i_u, i_{u+r}) \in \operatorname{Inv} \sigma. \tag{840}$$

[*Proof of (840):* The sequence $(i_1, i_2, i_3, \ldots)$ is periodic with period k, and its first k entries $i_1, i_2, \ldots, i_k$ are distinct. Hence, each entry of this sequence repeats itself each k steps, but not more often. Hence, every integer $u \geq 1$ satisfies $i_{u+r} \neq i_u$ (since $r \in \{1, 2, \ldots, k-1\}$). In other words, every $u \geq 1$ satisfies

$$i_{u+r} - i_u \neq 0. \tag{841}$$

The k-tuple $(i_{1+r} - i_1, i_{2+r} - i_2, \ldots, i_{k+r} - i_k)$ contains at least one positive entry[418],

---

[417] *Proof.* Every $u \in \{1, 2, 3, \ldots\}$ satisfies

$$\begin{aligned} i_u + i_{u+1} + \cdots + i_{u+k-1} &= \underbrace{i_u}_{\substack{= i_{u+k} \\ \text{(by (837))}}} + (i_{u+1} + i_{u+2} + \cdots + i_{u+k-1}) \\ &= i_{u+k} + (i_{u+1} + i_{u+2} + \cdots + i_{u+k-1}) = (i_{u+1} + i_{u+2} + \cdots + i_{u+k-1}) + i_{u+k} \\ &= i_{u+1} + i_{u+2} + \cdots + i_{u+k}. \end{aligned}$$

Thus, $i_1 + i_2 + \cdots + i_k = i_2 + i_3 + \cdots + i_{k+1} = i_3 + i_4 + \cdots + i_{k+2} = \cdots$, qed.

[418] *Proof.* Assume the contrary. Thus, the k-tuple $(i_{1+r} - i_1, i_{2+r} - i_2, \ldots, i_{k+r} - i_k)$ contains no positive entries. In other words, no $u \in \{1, 2, \ldots, k\}$ satisfies $i_{u+r} - i_u > 0$. In other words, every $u \in \{1, 2, \ldots, k\}$ satisfies $i_{u+r} - i_u \leq 0$. But (841) shows that every $u \in \{1, 2, \ldots, k\}$ satisfies $i_{u+r} - i_u \neq 0$. Thus, every $u \in \{1, 2, \ldots, k\}$ satisfies $i_{u+r} - i_u < 0$ (since $i_{u+r} - i_u \leq 0$ and

and at least one negative entry[419]. Hence, there exists at least one $u \geq 1$ such that $i_{u+r} - i_u > 0$ but $i_{(u+1)+r} - i_{u+1} < 0$ [420]. Consider this u. We have $i_u < i_{u+r}$ (since $i_{u+r} - i_u > 0$), so that $1 \leq i_u < i_{u+r} \leq n$. Also, (839) (applied to $p = u$) yields $\sigma(i_u) = i_{u+1}$. Moreover, (839) (applied to $p = u + r$) yields $\sigma(i_{u+r}) = i_{u+r+1} = i_{(u+1)+r}$. Hence,

$$\sigma(i_u) = i_{u+1} > i_{(u+1)+r} \qquad \left(\text{since } i_{(u+1)+r} - i_{u+1} < 0\right)$$
$$= \sigma(i_{u+r}).$$

So we know that (i_u, i_{u+r}) is a pair of integers satisfying $1 \leq i_u < i_{u+r} \leq n$ and $\sigma(i_u) > \sigma(i_{u+r})$. In other words, (i_u, i_{u+r}) is an inversion of σ. In other words, $(i_u, i_{u+r}) \in \operatorname{Inv} \sigma$. Thus, we have found a $u \geq 1$ such that $(i_u, i_{u+r}) \in \operatorname{Inv} \sigma$. This proves (840).]

$i_{u+r} - i_u \neq 0$). Hence,

$$\sum_{u=1}^{k} \underbrace{(i_{u+r} - i_u)}_{<0} < \sum_{u=1}^{k} 0 = 0.$$

But this contradicts

$$\sum_{u=1}^{k} \left(\underbrace{i_{u+r}}_{= i_{r+u}} - i_u \right) = \sum_{u=1}^{k} (i_{r+u} - i_u) = \left(\sum_{u=1}^{k} i_{r+u} \right) - \left(\sum_{u=1}^{k} i_u \right)$$
$$= (i_{r+1} + i_{r+2} + \cdots + i_{r+k}) - (i_1 + i_2 + \cdots + i_k) = 0 \qquad \text{(by (838))}.$$

This contradiction shows that our assumption was wrong, qed.

[419]This is proven similarly.

[420]*Proof.* Assume the contrary. Then, there exists no $u \geq 1$ such that $i_{u+r} - i_u > 0$ but $i_{(u+1)+r} - i_{u+1} < 0$. Hence,

$$\text{every } u \geq 1 \text{ satisfying } i_{u+r} - i_u > 0 \text{ must satisfy } i_{(u+1)+r} - i_{u+1} \geq 0. \tag{842}$$

Therefore,

$$\text{every } u \geq 1 \text{ satisfying } i_{u+r} - i_u > 0 \text{ must satisfy } i_{(u+1)+r} - i_{u+1} > 0 \tag{843}$$

(because (842) shows that $i_{(u+1)+r} - i_{u+1} \geq 0$; but combining this with $i_{(u+1)+r} - i_{u+1} \neq 0$ (which follows from (841), applied to $u + 1$ instead of u), we obtain $i_{(u+1)+r} - i_{u+1} > 0$).

But there exists some $v \in \{1, 2, \ldots, k\}$ satisfying $i_{v+r} - i_v > 0$ (since the k-tuple $(i_{1+r} - i_1, i_{2+r} - i_2, \ldots, i_{k+r} - i_k)$ contains at least one positive entry). Consider this v. Then, we have $i_{v+r} - i_v > 0$, therefore $i_{(v+1)+r} - i_{v+1} > 0$ (by (843), applied to $u = v$), therefore $i_{(v+2)+r} - i_{v+2} > 0$ (by (843), applied to $u = v + 1$), therefore $i_{(v+3)+r} - i_{v+3} > 0$ (by (843), applied to $u = v + 2$), and so on. Altogether, we thus obtain

$$i_{h+r} - i_h > 0 \qquad \text{for every } h \geq v.$$

In other words, $i_h < i_{h+r}$ for every $h \geq v$. Hence, $i_v < i_{v+r} < i_{v+2r} < i_{v+3r} < \cdots$. Thus, the numbers $i_v, i_{v+r}, i_{v+2r}, i_{v+3r}, \ldots$ are pairwise distinct; hence, the sequence $(i_1, i_2, i_3, \ldots)$ contains infinitely many distinct entries. But this contradicts the fact that this sequence is periodic. This contradiction proves that our assumption was wrong, qed.

Now, let us forget that we fixed r. We have shown that, for every $r \in \{1, 2, \ldots, k-1\}$, there exists some $u \geq 1$ such that $(i_u, i_{u+r}) \in \operatorname{Inv} \sigma$. Let us denote this u by u_r. Therefore, for every $r \in \{1, 2, \ldots, k-1\}$, we have found a $u_r \geq 1$ such that $(i_{u_r}, i_{u_r+r}) \in \operatorname{Inv} \sigma$. The $k-1$ pairs

$$\left(i_{u_1}, i_{u_1+1}\right), \left(i_{u_2}, i_{u_2+2}\right), \ldots, \left(i_{u_{k-1}}, i_{u_{k-1}+(k-1)}\right)$$

are pairwise distinct[421], and all of them belong to $\operatorname{Inv} \sigma$. Hence, the set $\operatorname{Inv} \sigma$ has at least $k-1$ elements. In other words, $|\operatorname{Inv} \sigma| \geq k-1$. Thus, $\ell(\sigma) = |\operatorname{Inv} \sigma| \geq k-1$. Since $\sigma = \operatorname{cyc}_{i_1, i_2, \ldots, i_k}$, this rewrites as $\ell\left(\operatorname{cyc}_{i_1, i_2, \ldots, i_k}\right) \geq k-1$. This solves Exercise 5.17 **(c)**.

(d) *First solution to Exercise 5.17 **(d)**:* Let $i_1, i_2, \ldots, i_k$ be k distinct elements of $[n]$. Hence, Proposition 5.6 **(c)** (applied to $(p_1, p_2, \ldots, p_k) = (i_1, i_2, \ldots, i_k)$) yields that there exists a permutation $\sigma \in S_n$ such that $(i_1, i_2, \ldots, i_k) = (\sigma(1), \sigma(2), \ldots, \sigma(k))$. Consider such a σ.

Exercise 5.17 **(b)** yields $\ell\left(\operatorname{cyc}_{1,2,\ldots,k}\right) = k-1$. But the definition of $(-1)^{\operatorname{cyc}_{1,2,\ldots,k}}$ yields $(-1)^{\operatorname{cyc}_{1,2,\ldots,k}} = (-1)^{\ell\left(\operatorname{cyc}_{1,2,\ldots,k}\right)} = (-1)^{k-1}$ (since $\ell\left(\operatorname{cyc}_{1,2,\ldots,k}\right) = k-1$).

Exercise 5.17 **(a)** (applied to $1, 2, \ldots, k$ instead of $i_1, i_2, \ldots, i_k$) yields

$$\sigma \circ \operatorname{cyc}_{1,2,\ldots,k} \circ \sigma^{-1} = \operatorname{cyc}_{\sigma(1), \sigma(2), \ldots, \sigma(k)} = \operatorname{cyc}_{i_1, i_2, \ldots, i_k}$$

(since $(\sigma(1), \sigma(2), \ldots, \sigma(k)) = (i_1, i_2, \ldots, i_k)$). Hence,

$$\underbrace{\operatorname{cyc}_{i_1, i_2, \ldots, i_k}}_{= \sigma \circ \operatorname{cyc}_{1,2,\ldots,k} \circ \sigma^{-1}} \circ \sigma = \sigma \circ \operatorname{cyc}_{1,2,\ldots,k} \circ \underbrace{\sigma^{-1} \circ \sigma}_{= \operatorname{id}} = \sigma \circ \operatorname{cyc}_{1,2,\ldots,k}.$$

Thus,

$$\begin{aligned} (-1)^{\operatorname{cyc}_{i_1, i_2, \ldots, i_k} \circ \sigma} = (-1)^{\sigma \circ \operatorname{cyc}_{1,2,\ldots,k}} &= (-1)^{\sigma} \cdot \underbrace{(-1)^{\operatorname{cyc}_{1,2,\ldots,k}}}_{= (-1)^{k-1}} \\ &\qquad \left(\text{by (315), applied to } \tau = \operatorname{cyc}_{1,2,\ldots,k}\right) \\ &= (-1)^{\sigma} \cdot (-1)^{k-1}. \end{aligned}$$

[421] *Proof.* Assume the contrary. Then, there exist two distinct elements x and y of $\{1, 2, \ldots, k-1\}$ such that $(i_{u_x}, i_{u_x+x}) = \left(i_{u_y}, i_{u_y+y}\right)$. Consider these x and y.

We have $(i_{u_x}, i_{u_x+x}) = \left(i_{u_y}, i_{u_y+y}\right)$. In other words, $i_{u_x} = i_{u_y}$ and $i_{u_x+x} = i_{u_y+y}$. Since the numbers $i_1, i_2, \ldots, i_k$ are distinct (and the sequence $(i_1, i_2, i_3, \ldots)$ consists of these numbers, repeated over and over), we obtain $u_x \equiv u_y \mod k$ from $i_{u_x} = i_{u_y}$, and we obtain $u_x + x \equiv u_y + y \mod k$ from $i_{u_x+x} = i_{u_y+y}$. Subtracting the congruence $u_x \equiv u_y \mod k$ from the congruence $u_x + x \equiv u_y + y \mod k$, we obtain $x \equiv y \mod k$. In light of $x, y \in \{1, 2, \ldots, k-1\}$, this shows that $x = y$. But this contradicts the fact that x and y are distinct. This contradiction proves that our assumption was wrong, qed.

Compared with

$$\begin{aligned}(-1)^{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}\circ\sigma} &= (-1)^{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}} \cdot (-1)^{\sigma} \\ &\qquad \left(\text{by (315), applied to } \operatorname{cyc}_{i_1,i_2,\ldots,i_k} \text{ and } \sigma \text{ instead of } \sigma \text{ and } \tau\right) \\ &= (-1)^{\sigma} \cdot (-1)^{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}},\end{aligned}$$

this yields $(-1)^{\sigma} \cdot (-1)^{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}} = (-1)^{\sigma} \cdot (-1)^{k-1}$. We can cancel $(-1)^{\sigma}$ from this equality (since $(-1)^{\sigma} \in \{1,-1\}$ is a nonzero integer), and thus obtain $(-1)^{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}} = (-1)^{k-1}$. This solves Exercise 5.17 **(d)**.

Second solution to Exercise 5.17 ***(d)****:* The following solution to Exercise 5.17 **(d)** appears in various textbooks on abstract algebra.

Let $i_1, i_2, \ldots, i_k$ be k distinct elements of $[n]$. We have

$$(-1)^{t_{i_j,i_{j+1}}} = -1 \qquad \text{for each } j \in \{1,2,\ldots,k-1\}. \tag{844}$$

Indeed, this follows from Exercise 5.10 **(b)** (applied to i_j and i_{j+1} instead of i and j).

Exercise 5.16 yields

$$\operatorname{cyc}_{i_1,i_2,\ldots,i_k} = t_{i_1,i_2} \circ t_{i_2,i_3} \circ \cdots \circ t_{i_{k-1},i_k}.$$

Hence,

$$\begin{aligned}(-1)^{\operatorname{cyc}_{i_1,i_2,\ldots,i_k}} &= (-1)^{t_{i_1,i_2}\circ t_{i_2,i_3}\circ\cdots\circ t_{i_{k-1},i_k}} = (-1)^{t_{i_1,i_2}} \cdot (-1)^{t_{i_2,i_3}} \cdot \cdots \cdot (-1)^{t_{i_{k-1},i_k}} \\ &\qquad \left(\begin{array}{c}\text{by Proposition 5.28 (applied to } k-1 \\ \text{and } t_{i_j,i_{j+1}} \text{ instead of } k \text{ and } \sigma_j)\end{array}\right) \\ &= \prod_{j=1}^{k-1} \underbrace{(-1)^{t_{i_j,i_{j+1}}}}_{\substack{=-1 \\ \text{(by (844))}}} = \prod_{j=1}^{k-1} (-1) = (-1)^{k-1}.\end{aligned}$$

This solves Exercise 5.17 **(d)** again. □

7.57. Solution to Exercise 5.18

We shall now approach the solution of Exercise 5.18.

Throughout Section 7.57, we shall use the same notations that were used in Section 5.8. In particular, n will denote a fixed element of $\mathbb{N}$.

We begin by proving Proposition 5.46:

Proof of Proposition 5.46. Recall that the inversions of σ are defined to be the pairs (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. This definition can be rewritten as follows: The inversions of σ are the pairs $(i,j) \in [n]^2$ satisfying $i < j$ and $\sigma(i) > \sigma(j)$. Thus,

$$\begin{aligned}
&(\text{the number of inversions of } \sigma) \\
&= \left(\text{the number of all pairs } (i,j) \in [n]^2 \text{ satisfying } i < j \text{ and } \sigma(i) > \sigma(j)\right) \\
&= \sum_{i \in [n]} \underbrace{(\text{the number of all } j \in [n] \text{ satisfying } i < j \text{ and } \sigma(i) > \sigma(j))}_{\substack{=(\text{the number of all } j \in \{i+1,i+2,\ldots,n\} \text{ and } \sigma(i) > \sigma(j)) \\ (\text{because the } j \in [n] \text{ satisfying } i<j \text{ are precisely the } j \in \{i+1,i+2,\ldots,n\})}} \\
&= \sum_{i \in [n]} \underbrace{(\text{the number of all } j \in \{i+1,i+2,\ldots,n\} \text{ and } \sigma(i) > \sigma(j))}_{\substack{=\ell_i(\sigma) \\ (\text{by the definition of } \ell_i(\sigma))}} = \sum_{i \in [n]} \ell_i(\sigma) \\
&= \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma).
\end{aligned}$$

This proves Proposition 5.46. □

Proof of Proposition 5.47. Recall that $H = [n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0$.

For each $i \in \{1,2,\ldots,n\}$, we have

$$\ell_i(\sigma) \in [n-i]_0. \tag{845}$$

[*Proof of (845):* Let $i \in \{1,2,\ldots,n\}$. Thus, $i \in \{1,2,\ldots,n\} = [n]$. Recall that $\ell_i(\sigma)$ is the number of all $j \in \{i+1,i+2,\ldots,n\}$ such that $\sigma(i) > \sigma(j)$ (by the definition of $\ell_i(\sigma)$). Thus,

$$\begin{aligned}
\ell_i(\sigma) &= \Bigg|\underbrace{\{j \in \{i+1,i+2,\ldots,n\} \mid \sigma(i) > \sigma(j)\}}_{\subseteq \{i+1,i+2,\ldots,n\}}\Bigg| \\
&\leq |\{i+1,i+2,\ldots,n\}| = n-i \qquad (\text{since } i \leq n).
\end{aligned}$$

Hence, $\ell_i(\sigma) \in \{0,1,\ldots,n-i\}$ (since $\ell_i(\sigma) \in \mathbb{N}$). But the definition of $[n-i]_0$ yields $[n-i]_0 = \{0,1,\ldots,n-i\}$. Hence, $\ell_i(\sigma) \in \{0,1,\ldots,n-i\} = [n-i]_0$. This proves (845).]

From (845), we know that $\ell_i(\sigma) \in [n-i]_0$ for each $i \in \{1,2,\ldots,n\}$. In other words,

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \in [n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0.$$

In view of $H = [n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0$, this rewrites as $(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \in H$. This proves Proposition 5.47. □

Proof of Lemma 5.48. **(a)** We know that $\ell_i(\sigma)$ is the number of all $j \in \{i+1,i+2,\ldots,n\}$ such that $\sigma(i) > \sigma(j)$ (by the definition of $\ell_i(\sigma)$). Hence,

$$\begin{aligned}
\ell_i(\sigma) &= (\text{the number of all } j \in \{i+1,i+2,\ldots,n\} \text{ such that } \sigma(i) > \sigma(j)) \\
&= |\{j \in \{i+1,i+2,\ldots,n\} \mid \sigma(i) > \sigma(j)\}|.
\end{aligned} \tag{846}$$

Define a set A by

$$A = \{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}. \qquad (847)$$

Thus,

$$|A| = |\{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}| = \ell_i(\sigma) \qquad (848)$$

(by (846)).

Let B be the set $[\sigma(i) - 1] \setminus \sigma([i])$.

The map σ is a permutation of $[n]$ (since $\sigma \in S_n$), and thus is invertible, and therefore is injective.

For each $k \in A$, we have $\sigma(k) \in B$ [422]. Hence, we can define a map $\alpha : A \to B$ by

$$(\alpha(k) = \sigma(k) \qquad \text{for each } k \in A).$$

Consider this α.

On the other hand, for each $k \in B$, we have $\sigma^{-1}(k) \in A$ [423]. Hence, we can define a map $\beta : B \to A$ by

$$\left(\beta(k) = \sigma^{-1}(k) \qquad \text{for each } k \in B\right).$$

---

[422] *Proof.* Let $k \in A$. Thus, $k \in A = \{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}$. In other words, k is an element of $\{i+1, i+2, \ldots, n\}$ and satisfies $\sigma(i) > \sigma(k)$.

From $k \in \{i+1, i+2, \ldots, n\} \subseteq [n]$, we conclude that $\sigma(k)$ is well-defined. Also, $\sigma(k) < \sigma(i)$ (since $\sigma(i) > \sigma(k)$), so that $\sigma(k) \leq \sigma(i) - 1$ (since $\sigma(k)$ and $\sigma(i)$ are integers). Thus, $\sigma(k) \in [\sigma(i) - 1]$.

Next, let us prove that $\sigma(k) \notin \sigma([i])$.

Indeed, assume the contrary (for the sake of contradiction). Hence, $\sigma(k) \in \sigma([i])$. In other words, there exists some $j \in [i]$ such that $\sigma(k) = \sigma(j)$. Consider this j. From $\sigma(k) = \sigma(j)$, we obtain $k = j$ (since the map σ is injective). Hence, $k = j \in [i]$. But $k \in \{i+1, i+2, \ldots, n\} = [n] \setminus [i]$, so that $k \notin [i]$. This contradicts $k \in [i]$. This contradiction shows that our assumption was false. Hence, $\sigma(k) \notin \sigma([i])$ is proven.

Combining $\sigma(k) \in [\sigma(i) - 1]$ with $\sigma(k) \notin \sigma([i])$, we obtain $\sigma(k) \in [\sigma(i) - 1] \setminus \sigma([i]) = B$. Qed.

[423] *Proof.* Let $k \in B$. Thus, $k \in B = [\sigma(i) - 1] \setminus \sigma([i])$. In other words, $k \in [\sigma(i) - 1]$ and $k \notin \sigma([i])$.

From $k \in [\sigma(i) - 1]$, we obtain $1 \leq k \leq \sigma(i) - 1$. Also, $k \in [\sigma(i) - 1] \subseteq [n]$, so that $\sigma^{-1}(k)$ is a well-defined element of $[n]$.

We have $\sigma\left(\sigma^{-1}(k)\right) = k \leq \sigma(i) - 1 < \sigma(i)$. In other words, $\sigma(i) > \sigma\left(\sigma^{-1}(k)\right)$.

Next, we claim that $\sigma^{-1}(k) \in \{i+1, i+2, \ldots, n\}$. Indeed, assume the contrary (for the sake of contradiction). Thus, $\sigma^{-1}(k) \notin \{i+1, i+2, \ldots, n\}$. Combining this with $\sigma^{-1}(k) \in [n]$, we obtain

$$\sigma^{-1}(k) \in [n] \setminus \{i+1, i+2, \ldots, n\} = [i].$$

Hence, $k = \sigma\left(\underbrace{\sigma^{-1}(k)}_{\in [i]}\right) \in \sigma([i])$, which contradicts $k \notin \sigma([i])$. This contradiction shows that our assumption was false. Thus, $\sigma^{-1}(k) \in \{i+1, i+2, \ldots, n\}$ is proven.

Now, we know that $\sigma^{-1}(k) \in \{i+1, i+2, \ldots, n\}$ and $\sigma(i) > \sigma\left(\sigma^{-1}(k)\right)$. In other words, $\sigma^{-1}(k)$ is a $j \in \{i+1, i+2, \ldots, n\}$ satisfying $\sigma(i) > \sigma(j)$. In other words,

$$\sigma^{-1}(k) \in \{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}.$$

In view of (847), this rewrites as $\sigma^{-1}(k) \in A$. Qed.

Consider this β.

The maps α and β are mutually inverse (since α is a restriction of σ, whereas β is a restriction of σ^{-1}), and therefore are bijections. Hence, there is a bijection from A to B (namely, α). Thus, $|A| = |B|$.

But (848) yields

$$\ell_i(\sigma) = |A| = |B| = |[\sigma(i) - 1] \setminus \sigma([i])|$$

(since $B = [\sigma(i) - 1] \setminus \sigma([i])$). This proves Lemma 5.48 **(a)**.

(b) If we had $\sigma(i) \in [\sigma(i) - 1]$, then we would have $\sigma(i) \leq \sigma(i) - 1 < \sigma(i)$, which would be absurd. Hence, we have $\sigma(i) \notin [\sigma(i) - 1]$.

But $[i] = \{i\} \cup [i-1]$. Hence,

$$\sigma\left(\underbrace{[i]}_{=\{i\}\cup[i-1]}\right) = \sigma(\{i\} \cup [i-1]) = \underbrace{\sigma(\{i\})}_{=\{\sigma(i)\}} \cup \sigma([i-1]) = \{\sigma(i)\} \cup \sigma([i-1]).$$

Thus,

$$\begin{aligned}
&[\sigma(i) - 1] \setminus \underbrace{\sigma([i])}_{=\{\sigma(i)\}\cup\sigma([i-1])} \\
&= [\sigma(i) - 1] \setminus (\{\sigma(i)\} \cup \sigma([i-1])) \\
&= \underbrace{([\sigma(i) - 1] \setminus \{\sigma(i)\})}_{\substack{=[\sigma(i)-1] \\ (\text{since } \sigma(i)\notin[\sigma(i)-1])}} \setminus \sigma([i-1]) = [\sigma(i) - 1] \setminus \sigma([i-1]).
\end{aligned}$$

Now, Lemma 5.48 **(a)** yields

$$\ell_i(\sigma) = \left|\underbrace{[\sigma(i) - 1] \setminus \sigma([i])}_{=[\sigma(i)-1]\setminus\sigma([i-1])}\right| = |[\sigma(i) - 1] \setminus \sigma([i-1])|.$$

This proves Lemma 5.48 **(b)**.

(c) The map σ is a permutation of $[n]$, and thus injective. Thus, Lemma 1.3 **(c)** (applied to $U = [n]$, $V = [n]$, $f = \sigma$ and $S = [i-1]$) yields $|\sigma([i-1])| = |[i-1]|$.

Now, Lemma 5.48 **(b)** yields

$$\begin{aligned}
\ell_i(\sigma) &= |[\sigma(i) - 1] \setminus \sigma([i-1])| \geq \underbrace{|[\sigma(i) - 1]|}_{\substack{=\sigma(i)-1 \\ (\text{since } |[k]|=k \\ \text{for every } k\in\mathbb{N})}} - \underbrace{|\sigma([i-1])|}_{\substack{=|[i-1]|=i-1 \\ (\text{since } |[k]|=k \\ \text{for every } k\in\mathbb{N})}} \\
&\qquad (\text{since any two finite sets } A \text{ and } B \text{ satisfies } |A \setminus B| \geq |A| - |B|) \\
&= (\sigma(i) - 1) - (i - 1) = \sigma(i) - i.
\end{aligned}$$

In other words, $\sigma(i) \leq i + \ell_i(\sigma)$. This proves Lemma 5.48 **(c)**. $\square$

Next, we state a further fact, which will be used in proving Proposition 5.50:

Lemma 7.107. Let $\sigma \in S_n$ and $\tau \in S_n$. Let $i \in [n]$. Assume that

$$\text{each } k \in [i-1] \text{ satisfies } \sigma(k) = \tau(k). \qquad (849)$$

Then:

(a) Each $k \in [i]$ satisfies $\sigma([k-1]) = \tau([k-1])$.

(b) We have $\ell_k(\sigma) = \ell_k(\tau)$ for each $k \in [i-1]$.

(c) Assume furthermore that $\sigma(i) < \tau(i)$. Then, $\ell_i(\sigma) < \ell_i(\tau)$.

Proof of Lemma 7.107. **(a)** Let $k \in [i]$. Thus, $k \leq i$.

Let $j \in [k-1]$. Thus, $j \leq \underbrace{k}_{\leq i} - 1 \leq i-1$, so that $j \in [i-1]$. Hence, (849) (applied to j instead of k) shows that $\sigma(j) = \tau(j)$.

Now, forget that we fixed j. We thus have shown that $\sigma(j) = \tau(j)$ for each $j \in [k-1]$. In other words,

$$(\sigma(1), \sigma(2), \ldots, \sigma(k-1)) = (\tau(1), \tau(2), \ldots, \tau(k-1)).$$

Thus,

$$\{\sigma(1), \sigma(2), \ldots, \sigma(k-1)\} = \{\tau(1), \tau(2), \ldots, \tau(k-1)\}.$$

Now,

$$\sigma\left(\underbrace{[k-1]}_{=\{1,2,\ldots,k-1\}}\right) = \sigma(\{1,2,\ldots,k-1\}) = \{\sigma(1), \sigma(2), \ldots, \sigma(k-1)\}$$
$$= \{\tau(1), \tau(2), \ldots, \tau(k-1)\} = \tau\left(\underbrace{\{1,2,\ldots,k-1\}}_{=[k-1]}\right) = \tau([k-1]).$$

This proves Lemma 7.107 **(a)**.

(b) Let $k \in [i-1]$. Then, Lemma 5.48 **(b)** (applied to k instead of i) yields $\ell_k(\sigma) = |[\sigma(k)-1] \setminus \sigma([k-1])|$. The same argument (applied to τ instead of σ) yields $\ell_k(\tau) = |[\tau(k)-1] \setminus \tau([k-1])|$.

But $k \in [i-1] \subseteq [i]$. Hence, Lemma 7.107 **(a)** yields $\sigma([k-1]) = \tau([k-1])$. Also, (849) yields $\sigma(k) = \tau(k)$. Hence,

$$\ell_k(\sigma) = \left|\left[\underbrace{\sigma(k)}_{=\tau(k)} - 1\right] \setminus \underbrace{\sigma([k-1])}_{=\tau([k-1])}\right| = |[\tau(k)-1] \setminus \tau([k-1])| = \ell_k(\tau).$$

This proves Lemma 7.107 **(b)**.

(c) We have $i \in [i]$. Hence, Lemma 7.107 **(a)** (applied to $k = i$) yields $\sigma([i-1]) = \tau([i-1])$.

Also, $\sigma(i) < \tau(i)$, so that $\sigma(i) - 1 < \tau(i) - 1$ and therefore $[\sigma(i) - 1] \subseteq [\tau(i) - 1]$.

From $\sigma(i) < \tau(i)$, we also obtain $\sigma(i) \leq \tau(i) - 1$ (since $\sigma(i)$ and $\tau(i)$ are integers), and thus $\sigma(i) \in [\tau(i) - 1]$.

Also, $\sigma(i) \notin \sigma([i-1])$. [*Proof:* Assume the contrary. Thus, $\sigma(i) \in \sigma([i-1])$. In other words, $\sigma(i) = \sigma(j)$ for some $j \in [i-1]$. Consider this j. From $\sigma(i) = \sigma(j)$, we obtain $i = j$ (since σ is injective), so that $i = j \in [i-1]$ and thus $i \leq i - 1 < i$. But this is absurd. Hence, we found a contradiction, so that $\sigma(i) \notin \sigma([i-1])$ is proven.]

If we had $\sigma(i) \in [\sigma(i) - 1]$, then we would have $\sigma(i) \leq \sigma(i) - 1 < \sigma(i)$, which is absurd. Hence, we have $\sigma(i) \notin [\sigma(i) - 1]$. Thus, also $\sigma(i) \notin [\sigma(i) - 1] \setminus \sigma([i-1])$.

Combining $\sigma(i) \in [\tau(i) - 1]$ with $\sigma(i) \notin \sigma([i-1])$, we obtain $\sigma(i) \in [\tau(i) - 1] \setminus \sigma([i-1])$.

Now,

$$\underbrace{[\sigma(i) - 1]}_{\subseteq [\tau(i) - 1]} \setminus \sigma([i-1]) \subseteq [\tau(i) - 1] \setminus \sigma([i-1]) . \tag{850}$$

Moreover, the set $[\tau(i) - 1] \setminus \sigma([i-1])$ contains $\sigma(i)$ (since $\sigma(i) \in [\tau(i) - 1] \setminus \sigma([i-1])$), but the set $[\sigma(i) - 1] \setminus \sigma([i-1])$ does not (since $\sigma(i) \notin [\sigma(i) - 1] \setminus \sigma([i-1])$). Thus, these two sets are distinct. In other words, $[\sigma(i) - 1] \setminus \sigma([i-1]) \neq [\tau(i) - 1] \setminus \sigma([i-1])$. Combining this with (850), we conclude that $[\sigma(i) - 1] \setminus \sigma([i-1])$ is a **proper** subset of $[\tau(i) - 1] \setminus \sigma([i-1])$.

But recall the following fundamental fact: If P is a finite set, and if Q is a proper subset of P, then $|Q| < |P|$. Applying this to $P = [\tau(i) - 1] \setminus \sigma([i-1])$ and $Q = [\sigma(i) - 1] \setminus \sigma([i-1])$, we conclude that

$$|[\sigma(i) - 1] \setminus \sigma([i-1])| < |[\tau(i) - 1] \setminus \sigma([i-1])|$$

(since $[\sigma(i) - 1] \setminus \sigma([i-1])$ is a **proper** subset of $[\tau(i) - 1] \setminus \sigma([i-1])$).

But Lemma 5.48 **(b)** yields $\ell_i(\sigma) = |[\sigma(i) - 1] \setminus \sigma([i-1])|$. The same argument (applied to τ instead of σ) yields $\ell_i(\tau) = |[\tau(i) - 1] \setminus \tau([i-1])|$. Hence,

$$\begin{aligned} \ell_i(\sigma) &= |[\sigma(i) - 1] \setminus \sigma([i-1])| \\ &< \left| [\tau(i) - 1] \setminus \underbrace{\sigma([i-1])}_{= \tau([i-1])} \right| = |[\tau(i) - 1] \setminus \tau([i-1])| = \ell_i(\tau) . \end{aligned}$$

This proves Lemma 7.107 **(c)**. ☐

We are now ready to prove Proposition 5.50:

Proof of Proposition 5.50. We have assumed that

$$(\sigma(1), \sigma(2), \ldots, \sigma(n)) <_{\text{lex}} (\tau(1), \tau(2), \ldots, \tau(n)) .$$

According to Definition 5.49, this means the following: There exists some $k \in [n]$ such that $\sigma(k) \neq \tau(k)$, and the **smallest** such k satisfies $\sigma(k) < \tau(k)$.

Let i be the smallest such k. Thus, i is an element of $[n]$ such that $\sigma(i) < \tau(i)$, but

$$\text{each } k \in [i-1] \text{ satisfies } \sigma(k) = \tau(k)$$

(since i is the **smallest** $k \in [n]$ such that $\sigma(k) \neq \tau(k)$).

Thus, Lemma 7.107 **(b)** shows that

$$\text{each } k \in [i-1] \text{ satisfies } \ell_k(\sigma) = \ell_k(\tau). \tag{851}$$

Furthermore, Lemma 7.107 **(c)** shows that $\ell_i(\sigma) < \ell_i(\tau)$ (since $\sigma(i) < \tau(i)$). Thus, $\ell_i(\sigma) \neq \ell_i(\tau)$. In other words, i is a $k \in [n]$ such that $\ell_k(\sigma) \neq \ell_k(\tau)$. Moreover, (851) shows that i is the **smallest** such k. Thus, the smallest $k \in [n]$ such that $\ell_k(\sigma) \neq \ell_k(\tau)$ satisfies $\ell_k(\sigma) < \ell_k(\tau)$ (because this k is i, and i satisfies $\ell_i(\sigma) < \ell_i(\tau)$).

Thus, we have shown that there exists some $k \in [n]$ such that $\ell_k(\sigma) \neq \ell_k(\tau)$, and the **smallest** such k satisfies $\ell_k(\sigma) < \ell_k(\tau)$. But this means precisely that

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) <_{\text{lex}} (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau))$$

(according to Definition 5.49). Hence, Proposition 5.50 is proven. □

Next, we prove a further lemma:

Lemma 7.108. Let $\sigma \in S_n$ and $\tau \in S_n$. Let $i \in [n]$. Assume that

$$\text{each } k \in [i-1] \text{ satisfies } \sigma(k) = \tau(k).$$

Assume furthermore that $\sigma(i) \neq \tau(i)$. Then, $\ell_i(\sigma) \neq \ell_i(\tau)$.

Proof of Lemma 7.108. We have $\sigma(i) \neq \tau(i)$. Hence, we are in one of the following two cases:

Case 1: We have $\sigma(i) < \tau(i)$.

Case 2: We have $\sigma(i) > \tau(i)$.

Let us first consider Case 1. In this case, we have $\sigma(i) < \tau(i)$. Hence, Lemma 7.107 **(c)** yields $\ell_i(\sigma) < \ell_i(\tau)$. Thus, $\ell_i(\sigma) \neq \ell_i(\tau)$. This proves Lemma 7.108 in Case 1.

Let us now consider Case 2. In this case, we have $\sigma(i) > \tau(i)$. In other words, $\tau(i) < \sigma(i)$.

But recall that each $k \in [i-1]$ satisfies $\sigma(k) = \tau(k)$. In other words, each $k \in [i-1]$ satisfies $\tau(k) = \sigma(k)$. Hence, Lemma 7.107 **(c)** (applied to τ and σ instead of σ and τ) yields $\ell_i(\tau) < \ell_i(\sigma)$. Thus, $\ell_i(\sigma) \neq \ell_i(\tau)$. Therefore, Lemma 7.108 is proven in Case 2.

We have now proven Lemma 7.108 in each of the two Cases 1 and 2. Hence, Lemma 7.108 always holds. □

Corollary 7.109. Let $\sigma \in S_n$ and $\tau \in S_n$ such that $L(\sigma) = L(\tau)$. Then, $\sigma = \tau$.

Proof of Corollary 7.109. Assume the contrary (for the sake of contradiction). Thus, $\sigma \neq \tau$. Hence, there exists some $k \in [n]$ satisfying $\sigma(k) \neq \tau(k)$.

Let i be the **smallest** such k. Thus, i is an element of $[n]$ and satisfies $\sigma(i) \neq \tau(i)$, but each $k \in [i-1]$ satisfies $\sigma(k) = \tau(k)$. Hence, Lemma 7.108 shows that $\ell_i(\sigma) \neq \ell_i(\tau)$.

But the definition of L yields $L(\sigma) = (\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma))$ and $L(\tau) = (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau))$. Thus,

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) = L(\sigma) = L(\tau) = (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau)).$$

In other words, each $j \in [n]$ satisfies $\ell_j(\sigma) = \ell_j(\tau)$. Applying this to $j = i$, we obtain $\ell_i(\sigma) = \ell_i(\tau)$. This contradicts $\ell_i(\sigma) \neq \ell_i(\tau)$. This contradiction completes our proof of Corollary 7.109. □

The next lemma is particularly obvious:

Lemma 7.110. We have $|H| = n!$.

Proof of Lemma 7.110. For each $k \in \mathbb{N}$, we have

$$|[k]_0| = k+1. \tag{852}$$

(This follows immediately from $[k]_0 = \{0, 1, \ldots, k\}$.)

But recall that $H = [n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0$. Hence,

$$\begin{aligned} |H| &= |[n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0| = |[n-1]_0| \cdot |[n-2]_0| \cdot \cdots \cdot |[n-n]_0| \\ &= \prod_{i=1}^{n} |[n-i]_0| = \prod_{k=0}^{n-1} \underbrace{|[k]_0|}_{\substack{=k+1 \\ \text{(by (852))}}} \qquad \left(\begin{array}{c}\text{here, we have substituted } k \text{ for } n-i \\ \text{in the product}\end{array}\right) \\ &= \prod_{k=0}^{n-1} (k+1) = \prod_{i=1}^{n} i \qquad (\text{here, we have substituted } i \text{ for } k+1 \text{ in the product}) \\ &= 1 \cdot 2 \cdot \cdots \cdot n = n!. \end{aligned}$$

This proves Lemma 7.110. □

Proof of Theorem 5.52. If $\sigma \in S_n$ and $\tau \in S_n$ are such that $L(\sigma) = L(\tau)$, then $\sigma = \tau$ (by Corollary 7.109). In other words, the map L is injective.

Lemma 7.110 shows that $|H| = n!$. But Corollary 7.81 shows that $|S_n| = n!$. Thus, $|S_n| = n! \geq n! = |H|$. Hence, Lemma 1.5 (applied to $U = S_n$, $V = H$ and $f = L$) shows that we have the following logical equivalence:

$$(L \text{ is injective}) \iff (L \text{ is bijective}).$$

Hence, L is bijective (since L is injective). In other words, the map $L : S_n \to H$ is a bijection. This proves Theorem 5.52. □

Finally, it remains to prove Corollary 5.53. We will need the following fact about products of polynomials:

Lemma 7.111. For every $i \in \{1, 2, \ldots, n\}$, let Z_i be a finite set. For every $i \in \{1, 2, \ldots, n\}$ and every $k \in Z_i$, let $p_{i,k}$ be a polynomial in x with rational coefficients. Then,

$$\prod_{i=1}^{n} \sum_{k \in Z_i} p_{i,k} = \sum_{(k_1, k_2, \ldots, k_n) \in Z_1 \times Z_2 \times \cdots \times Z_n} \prod_{i=1}^{n} p_{i,k_i}.$$

We defer the proof of Lemma 7.111 to a later section: namely, Lemma 7.111 is the particular case of Lemma 7.160 (which is proven below) obtained when $\mathbb{K}$ is the ring $\mathbb{Q}[x]$ (the ring of all polynomials in x with rational coefficients).

Proof of Corollary 5.53. Each $m \in \mathbb{N}$ satisfies $[m]_0 = \{0, 1, \ldots, m\}$ (by the definition of $[m]_0$) and thus

$$\sum_{k \in [m]_0} x^k = \underbrace{\sum_{k \in \{0,1,\ldots,m\}}}_{= \sum_{k=0}^{m}} x^k = \sum_{k=0}^{m} x^k. \tag{853}$$

Each $\sigma \in S_n$ satisfies

$$\ell(\sigma) = \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma)$$

(by Proposition 5.46) and therefore

$$x^{\ell(\sigma)} = x^{\ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma)} = x^{\ell_1(\sigma)} x^{\ell_2(\sigma)} \cdots x^{\ell_n(\sigma)} = \prod_{i=1}^{n} x^{\ell_i(\sigma)}. \tag{854}$$

Theorem 5.52 shows that the map L is a bijection. But the map $L : S_n \to H$ is defined by

$$\left(L(\sigma) = (\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \qquad \text{for each } \sigma \in S_n\right).$$

Thus, L is the map

$$S_n \to H, \qquad \sigma \mapsto (\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)).$$

Hence, the map

$$S_n \to H, \qquad \sigma \mapsto (\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma))$$

is a bijection (since L is a bijection).

We have

$$\sum_{w\in S_n} x^{\ell(w)} = \sum_{\sigma\in S_n} \underbrace{x^{\ell(\sigma)}}_{\substack{=\prod_{i=1}^{n} x^{\ell_i(\sigma)} \\ \text{(by (854))}}} \qquad \text{(here, we have renamed the summation index } w \text{ as } \sigma\text{)}$$
$$= \sum_{\sigma\in S_n} \prod_{i=1}^{n} x^{\ell_i(\sigma)} = \sum_{(k_1,k_2,\ldots,k_n)\in H} \prod_{i=1}^{n} x^{k_i}$$
$$\left(\begin{array}{c}\text{here, we have substituted } (k_1,k_2,\ldots,k_n) \\ \text{for } (\ell_1(\sigma),\ell_2(\sigma),\ldots,\ell_n(\sigma)) \text{ in the sum (since the map} \\ S_n\to H, \qquad \sigma\mapsto(\ell_1(\sigma),\ell_2(\sigma),\ldots,\ell_n(\sigma)) \\ \text{is a bijection)}\end{array}\right)$$
$$= \sum_{(k_1,k_2,\ldots,k_n)\in[n-1]_0\times[n-2]_0\times\cdots\times[n-n]_0} \prod_{i=1}^{n} x^{k_i} \tag{855}$$

(since $H=[n-1]_0\times[n-2]_0\times\cdots\times[n-n]_0$).

But Lemma 7.111 (applied to $Z_i=[n-i]_0$ and $p_{i,k}=x^k$) yields

$$\prod_{i=1}^{n}\sum_{k\in[n-i]_0} x^k = \sum_{(k_1,k_2,\ldots,k_n)\in[n-1]_0\times[n-2]_0\times\cdots\times[n-n]_0} \prod_{i=1}^{n} x^{k_i}.$$

Comparing this with (855), we obtain

$$\sum_{w\in S_n} x^{\ell(w)} = \prod_{i=1}^{n}\sum_{k\in[n-i]_0} x^k = \prod_{m=0}^{n-1}\underbrace{\sum_{k\in[m]_0} x^k}_{\substack{=\sum_{k=0}^{m} x^k \\ \text{(by (853))}}} \qquad \left(\begin{array}{c}\text{here, we have substituted } m \\ \text{for } n-i \text{ in the product}\end{array}\right)$$
$$= \prod_{m=0}^{n-1}\underbrace{\sum_{k=0}^{m} x^k}_{=1+x+x^2+\cdots+x^m} = \prod_{m=0}^{n-1}\left(1+x+x^2+\cdots+x^m\right)$$
$$= 1(1+x)\left(1+x+x^2\right)\cdots\left(1+x+x^2+\cdots+x^{n-1}\right)$$
$$= (1+x)\left(1+x+x^2\right)\cdots\left(1+x+x^2+\cdots+x^{n-1}\right).$$

This proves Corollary 5.53. □

7.58. Solution to Exercise 5.19

Throughout Section 7.58, we shall use the same notations that were in use throughout Section 5.8. We shall furthermore use the notation from Definition 3.48. We need several lemmas to prepare for the solution of Exercise 5.19:

Lemma 7.112. Let u and v be two integers satisfying $u \neq v$. Then,

$$[u > v] = 1 - [u < v] \tag{856}$$

and

$$[v > u] = [u < v] . \tag{857}$$

Proof of Lemma 7.112. The statement $u = v$ is false (since $u \neq v$). We have the following chain of logical equivalences:

$$(\text{not } u < v) \iff (u \geq v) \iff (u > v \text{ or } u = v) \iff (u > v)$$
$$(\text{since the statement } u = v \text{ is false}) .$$

Thus, $(\text{not } u < v)$ and $(u > v)$ are two equivalent logical statements. Hence, Exercise 3.12 **(a)** (applied to $\mathcal{A} = (\text{not } u < v)$ and $\mathcal{B} = (u > v)$) yields $[\text{not } u < v] = [u > v]$. Hence,

$$[u > v] = [\text{not } u < v] = 1 - [u < v]$$

(by Exercise 3.12 **(b)** (applied to $\mathcal{A} = (u < v)$)). This proves (856). Thus, it remains to prove (857).

Clearly, $(v > u)$ and $(u < v)$ are two equivalent logical statements. Hence, Exercise 3.12 **(a)** (applied to $\mathcal{A} = (v > u)$ and $\mathcal{B} = (u < v)$) yields $[v > u] = [u < v]$. Thus, (857) is proven. This completes the proof of Lemma 7.112. ☐

Lemma 7.113. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Let $i \in [n]$.

(a) We have $\sum\limits_{j \in [n]} [i < j \text{ and } \sigma(i) > \sigma(j)] = \ell_i(\sigma)$.

(b) We have $\sum\limits_{j \in [n]} [i < j] [\sigma(i) > \sigma(j)] = \ell_i(\sigma)$.

(c) We have $\sum\limits_{j \in [n]} [i < j] (1 - [\sigma(i) < \sigma(j)]) = \ell_i(\sigma)$.

Proof of Lemma 7.113. **(a)** We have

$$\begin{aligned}
&\sum_{j\in[n]} [i<j \text{ and } \sigma(i)>\sigma(j)] \\
&= \sum_{\substack{j\in[n];\\ i<j \text{ and } \sigma(i)>\sigma(j)}} \underbrace{[i<j \text{ and } \sigma(i)>\sigma(j)]}_{\substack{=1\\ (\text{since } i<j \text{ and } \sigma(i)>\sigma(j))}} \\
&\qquad + \sum_{\substack{j\in[n];\\ \text{not } (i<j \text{ and } \sigma(i)>\sigma(j))}} \underbrace{[i<j \text{ and } \sigma(i)>\sigma(j)]}_{\substack{=0\\ (\text{since we don't have } (i<j \text{ and } \sigma(i)>\sigma(j)))}} \\
&= \sum_{\substack{j\in[n];\\ i<j \text{ and } \sigma(i)>\sigma(j)}} 1 + \underbrace{\sum_{\substack{j\in[n];\\ \text{not } (i<j \text{ and } \sigma(i)>\sigma(j))}} 0}_{=0} = \sum_{\substack{j\in[n];\\ i<j \text{ and } \sigma(i)>\sigma(j)}} 1 \\
&= |\{j\in[n] \mid i<j \text{ and } \sigma(i)>\sigma(j)\}| \cdot 1 \\
&= |\{j\in[n] \mid i<j \text{ and } \sigma(i)>\sigma(j)\}| \\
&= (\text{the number of all } j\in[n] \text{ such that } i<j \text{ and } \sigma(i)>\sigma(j)) \\
&= (\text{the number of all } j\in[n] \text{ satisfying } i<j \text{ such that } \sigma(i)>\sigma(j)) \\
&= (\text{the number of all } j\in\{i+1,i+2,\ldots,n\} \text{ such that } \sigma(i)>\sigma(j)) \\
&\qquad (\text{since the } j\in[n] \text{ satisfying } i<j \text{ are precisely the } j\in\{i+1,i+2,\ldots,n\}) \\
&= \ell_i(\sigma)
\end{aligned}$$

(since $\ell_i(\sigma)$ was defined as the number of all $j \in \{i+1, i+2, \ldots, n\}$ such that $\sigma(i) > \sigma(j)$). This proves Lemma 7.113 **(a)**.

(b) Lemma 7.113 **(a)** yields

$$\ell_i(\sigma) = \sum_{j\in[n]} \underbrace{[i<j \text{ and } \sigma(i)>\sigma(j)]}_{\substack{=[(i<j)\wedge(\sigma(i)>\sigma(j))]\\ =[i<j][\sigma(i)>\sigma(j)]\\ (\text{by Exercise 3.12 } \textbf{(c)}\\ (\text{applied to } \mathcal{A}=(i<j) \text{ and } \mathcal{B}=(\sigma(i)>\sigma(j))))}} = \sum_{j\in[n]} [i<j][\sigma(i)>\sigma(j)].$$

This proves Lemma 7.113 **(b)**.

(c) We shall prove that each $j \in [n]$ satisfies

$$[i<j][\sigma(i)>\sigma(j)] = [i<j](1-[\sigma(i)<\sigma(j)]). \tag{858}$$

[*Proof of (858):* Let $j \in [n]$. If we don't have $i < j$, then we have $[i < j] = 0$. Hence, if we don't have $i < j$, then (858) rewrites as $0 = 0$, which is clearly true. Hence, for the rest of this proof of (858), we can WLOG assume that we do have $i < j$. Assume this.

Hence, $i \neq j$. But $\sigma \in S_n$. In other words, σ is a permutation of $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\}$). In other words, σ is a bijection $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. Hence, the map σ is bijective, therefore injective.

Hence, from $i \neq j$, we obtain $\sigma(i) \neq \sigma(j)$. Thus, (856) (applied to $u = \sigma(i)$ and $v = \sigma(j)$) yields $[\sigma(i) > \sigma(j)] = 1 - [\sigma(i) < \sigma(j)]$. Multiplying both sides of this equality with $[i < j]$, we obtain $[i < j][\sigma(i) > \sigma(j)] = [i < j](1 - [\sigma(i) < \sigma(j)])$. This proves (858).]

Now, Lemma 7.113 **(b)** yields

$$\ell_i(\sigma) = \sum_{j \in [n]} \underbrace{[i < j][\sigma(i) > \sigma(j)]}_{\substack{=[i<j](1-[\sigma(i)<\sigma(j)]) \\ \text{(by (858))}}} = \sum_{j \in [n]} [i < j](1 - [\sigma(i) < \sigma(j)]).$$

This proves Lemma 7.113 **(c)**. □

Lemma 7.114. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ and $\tau \in S_n$. Let $i \in [n]$ and $j \in [n]$. Then,

$$\begin{aligned} &[\tau(i) < \tau(j)](1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) + [i < j](1 - [\tau(i) < \tau(j)]) \\ &\qquad - [i < j](1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) \\ &= [j > i][\tau(i) > \tau(j)][\sigma(\tau(j)) > \sigma(\tau(i))] \\ &\qquad + [i > j][\tau(j) > \tau(i)][\sigma(\tau(i)) > \sigma(\tau(j))]. \end{aligned}$$

Proof of Lemma 7.114. We are in one of the following two cases:

Case 1: We have $i \neq j$.

Case 2: We have $i = j$.

Let us first consider Case 1. In this case, we have $i \neq j$.

Both τ and σ belong to S_n and thus are permutations of $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\}$). Hence, both τ and σ are bijective maps, and thus in particular are injective. From $i \neq j$, we obtain $\tau(i) \neq \tau(j)$ (since τ is injective) and therefore $\sigma(\tau(i)) \neq \sigma(\tau(j))$ (since σ is injective).

Define three integers a, b and c by

$$a = [i < j], \qquad b = [\tau(i) < \tau(j)] \qquad \text{and} \qquad c = [\sigma(\tau(i)) < \sigma(\tau(j))].$$

Comparing

$$\begin{aligned} &\underbrace{[\tau(i) < \tau(j)]}_{=b}\left(1 - \underbrace{[\sigma(\tau(i)) < \sigma(\tau(j))]}_{=c}\right) + \underbrace{[i < j]}_{=a}\left(1 - \underbrace{[\tau(i) < \tau(j)]}_{=b}\right) \\ &\qquad - \underbrace{[i < j]}_{=a}\left(1 - \underbrace{[\sigma(\tau(i)) < \sigma(\tau(j))]}_{=c}\right) \\ &= b(1-c) + a(1-b) - a(1-c) = b - bc - ab + ac \end{aligned}$$

with

$$\underbrace{[j>i]}_{\substack{=[i<j]\\ \text{(by (857)}\\ \text{(applied to } u=i \text{ and } v=j))}} \underbrace{[\tau(i)>\tau(j)]}_{\substack{=1-[\tau(i)<\tau(j)]\\ \text{(by (856)}\\ \text{(applied to } u=\tau(i) \text{ and } v=\tau(j)))}} \underbrace{[\sigma(\tau(j))>\sigma(\tau(i))]}_{\substack{=[\sigma(\tau(i))<\sigma(\tau(j))]\\ \text{(by (857)}\\ \text{(applied to } u=\sigma(\tau(i)) \text{ and } v=\sigma(\tau(j))))}}$$
$$+ \underbrace{[i>j]}_{\substack{=1-[i<j]\\ \text{(by (856)}\\ \text{(applied to } u=i \text{ and } v=j))}} \underbrace{[\tau(j)>\tau(i)]}_{\substack{=[\tau(i)<\tau(j)]\\ \text{(by (857)}\\ \text{(applied to } u=\tau(i) \text{ and } v=\tau(j)))}} \underbrace{[\sigma(\tau(i))>\sigma(\tau(j))]}_{\substack{=1-[\sigma(\tau(i))<\sigma(\tau(j))]\\ \text{(by (856)}\\ \text{(applied to } u=\sigma(\tau(i)) \text{ and } v=\sigma(\tau(j))))}}$$
$$= \underbrace{[i<j]}_{=a}\left(1-\underbrace{[\tau(i)<\tau(j)]}_{=b}\right)\underbrace{[\sigma(\tau(i))<\sigma(\tau(j))]}_{=c}$$
$$+\left(1-\underbrace{[i<j]}_{=a}\right)\underbrace{[\tau(i)<\tau(j)]}_{=b}\left(1-\underbrace{[\sigma(\tau(i))<\sigma(\tau(j))]}_{=c}\right)$$
$$= a(1-b)c+(1-a)b(1-c) = b-bc-ab+ac,$$

we obtain

$$\begin{aligned}
&[\tau(i)<\tau(j)](1-[\sigma(\tau(i))<\sigma(\tau(j))])+[i<j](1-[\tau(i)<\tau(j)]) \\
&\qquad -[i<j](1-[\sigma(\tau(i))<\sigma(\tau(j))]) \\
&=[j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))] \\
&\qquad +[i>j][\tau(j)>\tau(i)][\sigma(\tau(i))>\sigma(\tau(j))].
\end{aligned}$$

Thus, Lemma 7.114 is proven in Case 1.

Let us next consider Case 2. In this case, we have $i=j$. Thus, $i<j$ does not hold. Hence, $[i<j]=0$. Also, $j>i$ does not hold (since $j=i$). Hence, $[j>i]=0$. Also, $i>j$ does not hold (since $i=j$). Hence, $[i>j]=0$. Moreover, $\tau(i)<\tau(j)$ does not hold (since $\tau\left(\underbrace{i}_{=j}\right)=\tau(j)$). Hence, $[\tau(i)<\tau(j)]=0$. Comparing

$$\begin{aligned}
&\underbrace{[\tau(i)<\tau(j)]}_{=0}(1-[\sigma(\tau(i))<\sigma(\tau(j))])+\underbrace{[i<j]}_{=0}(1-[\tau(i)<\tau(j)]) \\
&\qquad -\underbrace{[i<j]}_{=0}(1-[\sigma(\tau(i))<\sigma(\tau(j))]) \\
&=0
\end{aligned}$$

with

$$\begin{aligned}
&\underbrace{[j > i]}_{=0} [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\
&\qquad + \underbrace{[i > j]}_{=0} [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] \\
&= 0,
\end{aligned}$$

we obtain

$$\begin{aligned}
&[\tau(i) < \tau(j)] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) + [i < j] (1 - [\tau(i) < \tau(j)]) \\
&\qquad - [i < j] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) \\
&= [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\
&\qquad + [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] .
\end{aligned}$$

Thus, Lemma 7.114 is proven in Case 2.

We have now proven Lemma 7.114 in each of the two Cases 1 and 2. Hence, Lemma 7.114 always holds. □

Let us now recall Definition 5.8.

Lemma 7.115. Let $n \in \mathbb{N}$. Let $k \in \{1, 2, \ldots, n-1\}$. Let $i \in [n]$. Then, $s_k(i) = i + [i = k] - [i = k+1]$.

Proof of Lemma 7.115. Recall that s_k is the permutation in S_n that switches k with $k+1$ but leaves all other numbers unchanged (by the definition of s_k). Thus, we have $s_k(k) = k+1$ and $s_k(k+1) = k$ and

$$(s_k(j) = j \qquad \text{for each } j \in \{1, 2, \ldots, n\} \setminus \{k, k+1\}) . \tag{859}$$

Now, we are in one of the following three cases:

Case 1: We have $i = k$.

Case 2: We have $i = k+1$.

Case 3: We have neither $i = k$ nor $i = k+1$.

Let us first consider Case 1. In this case, we have $i = k$. Thus, $s_k\left(\underbrace{i}_{=k}\right) = s_k(k) = k+1$. Comparing this with

$$\begin{aligned}
i + [i = k] - [i = k+1] &= k + \underbrace{[k = k]}_{\substack{=1 \\ (\text{since } k=k)}} - \underbrace{[k = k+1]}_{\substack{=0 \\ (\text{since } k=k+1 \text{ is false})}} \qquad (\text{since } i = k) \\
&= k + 1 - 0 = k+1,
\end{aligned}$$

we obtain $s_k(i) = i + [i = k] - [i = k+1]$. Hence, Lemma 7.115 is proven in Case 1.

An analogous argument proves Lemma 7.115 in Case 2.

Let us finally consider Case 3. In this case, we have neither $i = k$ nor $i = k + 1$. In other words, $i \notin \{k, k+1\}$. Combining $i \in [n] = \{1, 2, \ldots, n\}$ with $i \notin \{k, k+1\}$, we obtain $i \in \{1, 2, \ldots, n\} \setminus \{k, k+1\}$. Thus, (859) (applied to $j = i$) yields $s_k(i) = i$. Comparing this with

$$i + \underbrace{[i = k]}_{\substack{=0 \\ \text{(since we don't have } i=k)}} - \underbrace{[i = k+1]}_{\substack{=0 \\ \text{(since we don't have } i=k+1)}} = i + 0 - 0 = i,$$

we obtain $s_k(i) = i + [i = k] - [i = k+1]$. Hence, Lemma 7.115 is proven in Case 3.

We have now proven Lemma 7.115 in each of the three Cases 1, 2 and 3. Hence, Lemma 7.115 always holds. □

Lemma 7.116. Let $n \in \mathbb{N}$. Let $k \in \{1, 2, \ldots, n-1\}$. Let $i \in [n]$ and $j \in [n]$. Then:

(a) If $j > i$, then $[s_k(i) > s_k(j)] = [i = k][j = k+1]$.

(b) If $j \leq i$, then $[i = k][j = k+1] = 0$.

(c) We have

$$[j > i][s_k(i) > s_k(j)] = [i = k][j = k+1].$$

Proof of Lemma 7.116. The definition of s_k yields $s_k(k) = k+1$ and $s_k(k+1) = k$.

(a) Assume that $j > i$. Thus, $i < j$, so that $i \leq j - 1$ (since both i and j are integers). In other words, $i + 1 \leq j$.

Every statement $\mathcal{A}$ satisfies $[\mathcal{A}] \in \{0, 1\}$ and thus $[\mathcal{A}] \geq 0$ and $[\mathcal{A}] \leq 1$. These facts lead to $[i = k+1] \geq 0$, $[j = k] \geq 0$, $[j = k+1] \leq 1$ and $[i = k] \leq 1$.

We are in one of the following three cases:

Case 1: We have $i \neq k$.

Case 2: We have $j \neq k+1$.

Case 3: We have neither $i \neq k$ nor $j \neq k+1$.

(Of course, there can be overlap between Case 1 and Case 2.)

Let us first consider Case 1. In this case, we have $i \neq k$. In other words, we don't have $i = k$. Hence, $[i = k] = 0$. Now, Lemma 7.115 yields

$$s_k(i) = i + \underbrace{[i = k]}_{=0} - \underbrace{[i = k+1]}_{\geq 0} \leq i + 0 - 0 = i \leq j - 1.$$

Also, Lemma 7.115 (applied to j instead of i) yields

$$s_k(j) = j + \underbrace{[j = k]}_{\geq 0} - \underbrace{[j = k+1]}_{\leq 1} \geq j + 0 - 1 = j - 1.$$

Thus, $j - 1 \leq s_k(j)$, so that $s_k(i) \leq j - 1 \leq s_k(j)$. Thus, we cannot have $s_k(i) > s_k(j)$. Hence, $[s_k(i) > s_k(j)] = 0$. Comparing this with $\underbrace{[i = k]}_{=0}[j = k+1] = 0$, we obtain $[s_k(i) > s_k(j)] = [i = k][j = k+1]$. Hence, Lemma 7.116 **(a)** is proven in Case 1.

Let us next consider Case 2. In this case, we have $j \neq k+1$. In other words, we don't have $j = k+1$. Hence, $[j = k+1] = 0$. Now, Lemma 7.115 (applied to j instead of i) yields

$$s_k(j) = j + \underbrace{[j = k]}_{\geq 0} - \underbrace{[j = k+1]}_{=0} \geq j + 0 - 0 = j.$$

In other words, $j \leq s_k(j)$.

Also, Lemma 7.115 yields

$$s_k(i) = i + \underbrace{[i = k]}_{\leq 1} - \underbrace{[i = k+1]}_{\geq 0} \leq i + 1 - 0 = i + 1 \leq j \leq s_k(j).$$

Thus, we cannot have $s_k(i) > s_k(j)$. Hence, $[s_k(i) > s_k(j)] = 0$. Comparing this with $[i = k] \underbrace{[j = k+1]}_{=0} = 0$, we obtain $[s_k(i) > s_k(j)] = [i = k][j = k+1]$. Hence, Lemma 7.116 **(a)** is proven in Case 2.

Let us finally consider Case 3. In this case, we have neither $i \neq k$ nor $j \neq k+1$. Thus, we have both $i = k$ and $j = k+1$. Now, $s_k\left(\underbrace{i}_{=k}\right) = s_k(k) = k+1$ and $s_k\left(\underbrace{j}_{=k+1}\right) = s_k(k+1) = k$, so that $s_k(i) = k+1 > k = s_k(j)$. Hence, $[s_k(i) > s_k(j)] = 1$. Comparing this with $\underbrace{[i = k]}_{\substack{=1 \\ (\text{since } i=k)}} \underbrace{[j = k+1]}_{\substack{=1 \\ (\text{since } j=k+1)}} = 1$, we obtain $[s_k(i) > s_k(j)] = [i = k][j = k+1]$. Hence, Lemma 7.116 **(a)** is proven in Case 3.

We have now proven Lemma 7.116 **(a)** in each of the three Cases 1, 2 and 3. Hence, Lemma 7.116 **(a)** always holds.

(b) Assume that $j \leq i$. We are in one of the following two cases:

Case 1: We have $i \leq k$.

Case 2: We have $i > k$.

Let us first consider Case 1. In this case, we have $i \leq k$. Thus, $j \leq i \leq k < k+1$. Hence, we don't have $j = k+1$. Thus, $[j = k+1] = 0$. Thus, $[i = k] \underbrace{[j = k+1]}_{=0} = 0$. Hence, Lemma 7.116 **(b)** is proven in Case 1.

Let us now consider Case 2. In this case, we have $i > k$. Hence, we don't have $i = k$. Thus, $[i = k] = 0$. Thus, $\underbrace{[i = k]}_{=0} [j = k+1] = 0$. Hence, Lemma 7.116 **(b)** is proven in Case 2.

We have now proven Lemma 7.116 **(b)** in each of the two Cases 1 and 2. Thus, Lemma 7.116 **(b)** always holds.

(c) We are in one of the following two cases:

Case 1: We have $j > i$.

Case 2: We don't have $j > i$.

Let us first consider Case 1. In this case, we have $j > i$. Hence, $[j > i] = 1$. Hence, $\underbrace{[j > i]}_{=1} [s_k(i) > s_k(j)] = [s_k(i) > s_k(j)] = [i = k][j = k+1]$ (by Lemma 7.116 **(a)**). Thus, Lemma 7.116 **(c)** is proven in Case 1.

Let us next consider Case 2. In this case, we don't have $j > i$. Hence, $[j > i] = 0$. Also, $j \leq i$ (since we don't have $j > i$); thus, Lemma 7.116 **(b)** yields $[i = k][j = k+1] = 0$. Comparing this with $\underbrace{[j > i]}_{=0} [s_k(i) > s_k(j)] = 0$, we obtain $[j > i][s_k(i) > s_k(j)] = [i = k][j = k+1]$. Therefore, Lemma 7.116 **(c)** is proven in Case 2.

We have now proven Lemma 7.116 **(c)** in each of the two Cases 1 and 2. Thus, Lemma 7.116 **(c)** always holds. □

Lemma 7.117. Let P be a finite set. Let $p \in P$. Then, $\sum_{j \in P} [j = p] = 1$.

Proof of Lemma 7.117. We have $p \in P$. Thus, we can split off the addend for $j = p$ from the sum $\sum_{j \in P} [j = p]$. We thus obtain

$$\sum_{j \in P} [j = p] = \underbrace{[p = p]}_{\substack{=1 \\ (\text{since } p = p)}} + \sum_{\substack{j \in P; \\ j \neq p}} \underbrace{[j = p]}_{\substack{=0 \\ (\text{since we don't have } j = p \\ (\text{since } j \neq p))}} = 1 + \underbrace{\sum_{\substack{j \in P; \\ j \neq p}} 0}_{=0} = 1.$$

This proves Lemma 7.117. □

Solution to Exercise 5.19. We have $[n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$).

We have $\tau \in S_n$. In other words, τ is a permutation of $\{1, 2, \ldots, n\}$. In other words, τ is a permutation of $[n]$ (since $[n] = \{1, 2, \ldots, n\}$). In other words, τ is a bijection $[n] \to [n]$.

(a) Let $i \in [n]$. Then, $\tau(i) \in [n]$ (since τ is a bijection $[n] \to [n]$). Hence, Lemma 7.113 **(c)** (applied to $\tau(i)$ instead of i) yields

$$\sum_{j \in [n]} [\tau(i) < j] (1 - [\sigma(\tau(i)) < \sigma(j)]) = \ell_{\tau(i)}(\sigma).$$

Hence,

$$\begin{aligned} \ell_{\tau(i)}(\sigma) &= \sum_{j \in [n]} [\tau(i) < j] (1 - [\sigma(\tau(i)) < \sigma(j)]) \\ &= \sum_{j \in [n]} [\tau(i) < \tau(j)] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) \end{aligned} \tag{860}$$

(here, we have substituted $\tau(j)$ for j in the sum, since the map $\tau : [n] \to [n]$ is a bijection).

Also, Lemma 7.113 **(c)** (applied to $\sigma \circ \tau$ instead of σ) yields

$$\sum_{j \in [n]} [i < j] (1 - [(\sigma \circ \tau)(i) < (\sigma \circ \tau)(j)]) = \ell_i (\sigma \circ \tau).$$

Hence,

$$\begin{aligned} \ell_i (\sigma \circ \tau) &= \sum_{j \in [n]} [i < j] \left(1 - \left[\underbrace{(\sigma \circ \tau)(i)}_{=\sigma(\tau(i))} < \underbrace{(\sigma \circ \tau)(j)}_{=\sigma(\tau(j))}\right]\right) \\ &= \sum_{j \in [n]} [i < j] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]). \end{aligned} \tag{861}$$

Moreover, Lemma 7.113 **(c)** (applied to τ instead of σ) yields

$$\sum_{j \in [n]} [i < j] (1 - [\tau(i) < \tau(j)]) = \ell_i(\tau). \tag{862}$$

But Lemma 7.114 shows that every $j \in [n]$ satisfies

$$\begin{aligned} &[\tau(i) < \tau(j)] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) + [i < j] (1 - [\tau(i) < \tau(j)]) \\ &\qquad - [i < j] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) \\ &= [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\ &\qquad + [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))]. \end{aligned}$$

Summing up these equalities for all $j \in [n]$, we obtain

$$\begin{aligned} &\sum_{j \in [n]} \Big([\tau(i) < \tau(j)] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) + [i < j] (1 - [\tau(i) < \tau(j)]) \\ &\qquad\qquad - [i < j] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) \Big) \\ &= \sum_{j \in [n]} \Big([j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\ &\qquad\qquad + [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] \Big). \end{aligned} \tag{863}$$

Now,

$$\underbrace{\ell_{\tau(i)}(\sigma)}_{\substack{=\sum_{j\in[n]}[\tau(i)<\tau(j)](1-[\sigma(\tau(i))<\sigma(\tau(j))]) \\ \text{(by (860))}}} + \underbrace{\ell_i(\tau)}_{\substack{=\sum_{j\in[n]}[i<j](1-[\tau(i)<\tau(j)]) \\ \text{(by (862))}}} - \underbrace{\ell_i(\sigma\circ\tau)}_{\substack{=\sum_{j\in[n]}[i<j](1-[\sigma(\tau(i))<\sigma(\tau(j))]) \\ \text{(by (861))}}}$$
$$= \sum_{j\in[n]} [\tau(i) < \tau(j)] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) + \sum_{j\in[n]} [i < j] (1 - [\tau(i) < \tau(j)])$$
$$\qquad - \sum_{j\in[n]} [i < j] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))])$$
$$= \sum_{j\in[n]} \Big([\tau(i) < \tau(j)] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) + [i < j] (1 - [\tau(i) < \tau(j)])$$
$$\qquad - [i < j] (1 - [\sigma(\tau(i)) < \sigma(\tau(j))]) \Big)$$
$$= \sum_{j\in[n]} \Big([j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))]$$
$$\qquad + [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] \Big)$$
$$\qquad \text{(by (863))}$$
$$= \sum_{j\in[n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))]$$
$$\qquad + \sum_{j\in[n]} [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] .$$

This solves Exercise 5.19 **(a)**.

(b) We have $[n] = \{1, 2, \ldots, n\}$ and thus $\sum\limits_{i\in[n]} = \sum\limits_{i\in\{1,2,\ldots,n\}} = \sum\limits_{i=1}^{n}$ (an equality between summation signs). Now, Proposition 5.46 yields

$$\ell(\sigma) = \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma) = \underbrace{\sum_{i=1}^{n}}_{=\sum_{i\in[n]}} \ell_i(\sigma) = \sum_{i\in[n]} \ell_i(\sigma) = \sum_{i\in[n]} \ell_{\tau(i)}(\sigma)$$

(here, we have substituted $\tau(i)$ for i in the sum, since the map $\tau : [n] \to [n]$ is a bijection). Also, Proposition 5.46 (applied to τ instead of σ) yields

$$\ell(\tau) = \ell_1(\tau) + \ell_2(\tau) + \cdots + \ell_n(\tau) = \underbrace{\sum_{i=1}^{n}}_{=\sum_{i\in[n]}} \ell_i(\tau) = \sum_{i\in[n]} \ell_i(\tau) .$$

Moreover, Proposition 5.46 (applied to $\sigma \circ \tau$ instead of σ) yields

$$\ell(\sigma \circ \tau) = \ell_1(\sigma \circ \tau) + \ell_2(\sigma \circ \tau) + \cdots + \ell_n(\sigma \circ \tau) = \underbrace{\sum_{i=1}^{n}}_{= \sum_{i \in [n]}} \ell_i(\sigma \circ \tau) = \sum_{i \in [n]} \ell_i(\sigma \circ \tau).$$

Now,

$$\begin{aligned}
&\underbrace{\ell(\sigma)}_{= \sum_{i \in [n]} \ell_{\tau(i)}(\sigma)} + \underbrace{\ell(\tau)}_{= \sum_{i \in [n]} \ell_i(\tau)} - \underbrace{\ell(\sigma \circ \tau)}_{= \sum_{i \in [n]} \ell_i(\sigma \circ \tau)} \\
&= \sum_{i \in [n]} \ell_{\tau(i)}(\sigma) + \sum_{i \in [n]} \ell_i(\tau) - \sum_{i \in [n]} \ell_i(\sigma \circ \tau) \\
&= \sum_{i \in [n]} \underbrace{\left(\ell_{\tau(i)}(\sigma) + \ell_i(\tau) - \ell_i(\sigma \circ \tau)\right)}_{\substack{= \sum_{j \in [n]} [j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))] \\ + \sum_{j \in [n]} [i>j][\tau(j)>\tau(i)][\sigma(\tau(i))>\sigma(\tau(j))] \\ \text{(by Exercise 5.19 (a))}}} \\
&= \sum_{i \in [n]} \left(\sum_{j \in [n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \right. \\
&\qquad \left. + \sum_{j \in [n]} [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] \right) \\
&= \sum_{i \in [n]} \sum_{j \in [n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\
&\qquad + \sum_{i \in [n]} \sum_{j \in [n]} [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))].
\end{aligned} \tag{864}$$

But

$$\begin{aligned}
&\sum_{i \in [n]} \sum_{j \in [n]} [i > j] [\tau(j) > \tau(i)] [\sigma(\tau(i)) > \sigma(\tau(j))] \\
&= \underbrace{\sum_{j \in [n]} \sum_{i \in [n]}}_{= \sum_{i \in [n]} \sum_{j \in [n]}} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))] \\
&\qquad \left(\begin{array}{c} \text{here, we have renamed the summation indices } i \text{ and } j \\ \text{as } j \text{ and } i \text{, respectively} \end{array} \right) \\
&= \sum_{i \in [n]} \sum_{j \in [n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))].
\end{aligned}$$

Hence, (864) becomes

$$\begin{aligned}
&\ell(\sigma)+\ell(\tau)-\ell(\sigma\circ\tau)\\
&=\sum_{i\in[n]}\sum_{j\in[n]}[j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))]\\
&\qquad+\underbrace{\sum_{i\in[n]}\sum_{j\in[n]}[i>j][\tau(j)>\tau(i)][\sigma(\tau(i))>\sigma(\tau(j))]}_{=\sum_{i\in[n]}\sum_{j\in[n]}[j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))]}\\
&=\sum_{i\in[n]}\sum_{j\in[n]}[j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))]\\
&\qquad+\sum_{i\in[n]}\sum_{j\in[n]}[j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))]\\
&=2\sum_{i\in[n]}\sum_{j\in[n]}[j>i][\tau(i)>\tau(j)][\sigma(\tau(j))>\sigma(\tau(i))].
\end{aligned}$$

This solves Exercise 5.19 **(b)**.

(c) *Second solution to Exercise 5.2* ***(a)***. Forget that we fixed σ and τ.

Let $k\in\{1,2,\ldots,n-1\}$. Thus, $k+1\in\{2,3,\ldots,n\}\subseteq\{1,2,\ldots,n\}=[n]$. Also, $k\in\{1,2,\ldots,n-1\}\subseteq\{1,2,\ldots,n\}=[n]$.

Recall that s_k is the permutation in S_n that switches k with $k+1$ but leaves all other numbers unchanged (by the definition of s_k). Thus, we have $s_k(k)=k+1$ and $s_k(k+1)=k$ and

$$(s_k(j)=j\qquad\text{for each } j\in\{1,2,\ldots,n\}\setminus\{k,k+1\}).$$

Let us prove some auxiliary claims:

Claim 1: We have $\ell(s_k)=1$.

[*Proof of Claim 1:* Let $i\in[n]$. Lemma 7.113 **(b)** (applied to $\sigma=s_k$) yields

$$\sum_{j\in[n]}[i<j][s_k(i)>s_k(j)]=\ell_i(s_k).$$

Hence,

$$\begin{aligned}
\ell_i(s_k)&=\sum_{j\in[n]}\underbrace{[i<j]}_{\substack{=[j>i]\\ \text{(since the statement } i<j\\ \text{is equivalent to } j>i)}}[s_k(i)>s_k(j)]=\sum_{j\in[n]}\underbrace{[j>i][s_k(i)>s_k(j)]}_{\substack{=[i=k][j=k+1]\\ \text{(by Lemma 7.116 (c))}}}\\
&=\sum_{j\in[n]}[i=k][j=k+1]=[i=k]\underbrace{\sum_{j\in[n]}[j=k+1]}_{\substack{=1\\ \text{(by Lemma 7.117,}\\ \text{applied to } P=[n] \text{ and } p=k+1)}}\\
&=[i=k]. \qquad (865)
\end{aligned}$$

Now, forget that we fixed i. We thus have proven (865) for each $i \in [n]$.
Now, Proposition 5.46 (applied to $\sigma = s_k$) yields

$$\ell(s_k) = \ell_1(s_k) + \ell_2(s_k) + \cdots + \ell_n(s_k) = \sum_{i\in[n]} \underbrace{\ell_i(s_k)}_{\substack{=[i=k] \\ \text{(by (865))}}} = \sum_{i\in[n]} [i=k]$$
$$= \sum_{j\in[n]} [j=k] \qquad \text{(here, we have renamed the summation index } i \text{ as } j\text{)}$$
$$= 1 \qquad \text{(by Lemma 7.117, applied to } P = [n] \text{ and } p = k\text{)}.$$

This proves Claim 1.]

Claim 2: Let $\sigma \in S_n$. Then, $\ell(\sigma \circ s_k) = \ell(\sigma) + 1 - 2[\sigma(k) > \sigma(k+1)]$.

[*Proof of Claim 2:* Every $i \in [n]$ satisfies

$$\sum_{j\in[n]} \underbrace{[j>i][s_k(i) > s_k(j)]}_{\substack{=[i=k][j=k+1] \\ \text{(by Lemma 7.116 (c))}}} [\sigma(s_k(j)) > \sigma(s_k(i))]$$
$$= \sum_{j\in[n]} [i=k][j=k+1][\sigma(s_k(j)) > \sigma(s_k(i))]$$
$$= [i=k] \underbrace{[k+1=k+1]}_{\substack{=1 \\ \text{(since } k+1=k+1\text{)}}} \left[\sigma\left(\underbrace{s_k(k+1)}_{=k}\right) > \sigma(s_k(i))\right]$$
$$+ \sum_{\substack{j\in[n]; \\ j\neq k+1}} [i=k] \underbrace{[j=k+1]}_{\substack{=0 \\ \text{(since } j\neq k+1\text{)}}} [\sigma(s_k(j)) > \sigma(s_k(i))]$$
$$\left(\begin{array}{c}\text{here, we have split off the addend} \\ \text{for } j=k+1 \text{ from the sum (since } k+1\in[n]\text{)}\end{array}\right)$$
$$= [i=k][\sigma(k) > \sigma(s_k(i))] + \underbrace{\sum_{\substack{j\in[n]; \\ j\neq k+1}} [i=k]\, 0\, [\sigma(s_k(j)) > \sigma(s_k(i))]}_{=0}$$
$$= [i=k][\sigma(k) > \sigma(s_k(i))].$$

Summing up these equalities for all $i \in [n]$, we obtain

$$\begin{aligned}
&\sum_{i\in[n]} \sum_{j\in[n]} [j > i] [s_k(i) > s_k(j)] [\sigma(s_k(j)) > \sigma(s_k(i))] \\
&= \sum_{i\in[n]} [i = k] [\sigma(k) > \sigma(s_k(i))] \\
&= \underbrace{[k = k]}_{\substack{=1 \\ (\text{since } k=k)}} \left[\sigma(k) > \sigma\left(\underbrace{s_k(k)}_{=k+1}\right)\right] + \sum_{\substack{i\in[n]; \\ i\neq k}} \underbrace{[i = k]}_{\substack{=0 \\ (\text{since } i\neq k)}} [\sigma(k) > \sigma(s_k(i))] \\
&\qquad \text{(here, we have split off the addend for } i = k \text{ from the sum (since } k \in [n]\text{))} \\
&= [\sigma(k) > \sigma(k+1)] + \underbrace{\sum_{\substack{i\in[n]; \\ i\neq k}} 0 [\sigma(k) > \sigma(s_k(i))]}_{=0} = [\sigma(k) > \sigma(k+1)].
\end{aligned}$$

But Exercise 5.19 **(b)** (applied to $\tau = s_k$) yields

$$\begin{aligned}
&\ell(\sigma) + \ell(s_k) - \ell(\sigma \circ s_k) \\
&= 2 \underbrace{\sum_{i\in[n]} \sum_{j\in[n]} [j > i] [s_k(i) > s_k(j)] [\sigma(s_k(j)) > \sigma(s_k(i))]}_{=[\sigma(k)>\sigma(k+1)]} = 2 [\sigma(k) > \sigma(k+1)].
\end{aligned}$$

Solving this equation for $\ell(\sigma \circ s_k)$, we obtain

$$\begin{aligned}
\ell(\sigma \circ s_k) &= \ell(\sigma) + \underbrace{\ell(s_k)}_{\substack{=1 \\ (\text{by Claim 1})}} - 2 [\sigma(k) > \sigma(k+1)] \\
&= \ell(\sigma) + 1 - 2 [\sigma(k) > \sigma(k+1)].
\end{aligned}$$

This proves Claim 2.]

Claim 3: Let $\sigma \in S_n$. Then,

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases}.$$

[*Proof of Claim 3:* We have $\sigma \in S_n$. In other words, σ is a permutation of $\{1, 2, \ldots, n\}$. Thus, σ is a bijective map, and therefore an injective map. Hence, from $k \neq k+1$, we obtain $\sigma(k) \neq \sigma(k+1)$. Thus, the statement $(\sigma(k) \leq \sigma(k+1))$

is equivalent to the statement $(\sigma(k) < \sigma(k+1))$. Moreover, Claim 2 yields

$$\ell(\sigma \circ s_k) = \ell(\sigma) + 1 - 2 \underbrace{[\sigma(k) > \sigma(k+1)]}_{\substack{= \begin{cases} 0, & \text{if } \sigma(k) \leq \sigma(k+1); \\ 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases} \\ = \begin{cases} 0, & \text{if } \sigma(k) < \sigma(k+1); \\ 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases} \\ \text{(since the statement } (\sigma(k) \leq \sigma(k+1)) \text{ is equivalent} \\ \text{to the statement } (\sigma(k) < \sigma(k+1)))}}$$
$$= \ell(\sigma) + 1 - 2 \begin{cases} 0, & \text{if } \sigma(k) < \sigma(k+1); \\ 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases}$$
$$= \begin{cases} \ell(\sigma) + 1 - 2 \cdot 0, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) + 1 - 2 \cdot 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases}$$
$$= \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases}.$$

This proves Claim 3.]

Claim 4: Let $\sigma \in S_n$. Then,

$$\ell(s_k \circ \sigma) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1) \end{cases}.$$

[*Proof of Claim 4:* Recall that $s_i^2 = \mathrm{id}$ for each $i \in \{1, 2, \ldots, n-1\}$. Applying this to $i = k$, we obtain $s_k^2 = \mathrm{id}$; thus, $s_k \circ s_k = s_k^2 = \mathrm{id}$ and therefore $s_k^{-1} = s_k$.

Let us recall that $(\alpha \circ \beta)^{-1} = \beta^{-1} \circ \alpha^{-1}$ for any two permutations α and β in S_n. Applying this to $\alpha = \sigma^{-1}$ and $\beta = s_k$, we obtain $\left(\sigma^{-1} \circ s_k\right)^{-1} = \underbrace{s_k^{-1}}_{= s_k} \circ \underbrace{\left(\sigma^{-1}\right)^{-1}}_{= \sigma} = s_k \circ \sigma$. But Exercise 5.2 **(f)** yields $\ell(\sigma) = \ell\left(\sigma^{-1}\right)$. Also, Exercise 5.2 **(f)** (applied to $\sigma^{-1} \circ s_k$ instead of σ) yields $\ell\left(\sigma^{-1} \circ s_k\right) = \ell\left(\underbrace{\left(\sigma^{-1} \circ s_k\right)^{-1}}_{= s_k \circ \sigma}\right) = \ell(s_k \circ \sigma)$. But Claim 3 (applied to σ^{-1} instead of σ) yields

$$\ell\left(\sigma^{-1} \circ s_k\right) = \begin{cases} \ell\left(\sigma^{-1}\right) + 1, & \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1); \\ \ell\left(\sigma^{-1}\right) - 1, & \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1) \end{cases}.$$

Since $\ell\left(\sigma^{-1} \circ s_k\right) = \ell(s_k \circ \sigma)$ and $\ell\left(\sigma^{-1}\right) = \ell(\sigma)$, this equality rewrites as follows:

$$\ell(s_k \circ \sigma) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1) \end{cases}.$$

This proves Claim 4.]

Combining Claim 3 with Claim 4, we obtain the exact statement of Exercise 5.2 **(a)**. Thus, Exercise 5.2 **(a)** is solved again. This solves Exercise 5.19 **(c)**.

(d) *Second solution to Exercise 5.2* ***(b)****.* Let $\sigma \in S_n$ and $\tau \in S_n$. Exercise 5.19 **(b)** yields

$$\ell(\sigma) + \ell(\tau) - \ell(\sigma \circ \tau) = 2 \underbrace{\sum_{i \in [n]} \sum_{j \in [n]} [j > i] [\tau(i) > \tau(j)] [\sigma(\tau(j)) > \sigma(\tau(i))]}_{\substack{\text{This is an integer} \\ \text{(since the truth values } [j>i],\ [\tau(i)>\tau(j)],\ [\sigma(\tau(j))>\sigma(\tau(i))] \text{ are integers)}}} \equiv 0 \mod 2.$$

In other words, $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \mod 2$. Thus, Exercise 5.2 **(b)** is solved again. This solves Exercise 5.19 **(d)**.

(e) *Second solution to Exercise 5.2* ***(c)****.* Let $\sigma \in S_n$ and $\tau \in S_n$. Exercise 5.19 **(b)** yields

$$\ell(\sigma) + \ell(\tau) - \ell(\sigma \circ \tau) = 2 \sum_{i \in [n]} \sum_{j \in [n]} \underbrace{[j > i]}_{\geq 0} \underbrace{[\tau(i) > \tau(j)]}_{\geq 0} \underbrace{[\sigma(\tau(j)) > \sigma(\tau(i))]}_{\geq 0} \geq 0.$$

In other words, $\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau)$. Thus, Exercise 5.2 **(c)** is solved again. This solves Exercise 5.19 **(e)**. □

7.59. Solution to Exercise 5.20

Throughout Section 7.59, we shall use the same notations that were in use throughout Section 5.8. We shall furthermore use the notation from Definition 3.48. We need several lemmas to prepare for the solution of Exercise 5.20. We start with a lemma that restates the definition of $\ell_i(\sigma)$:

Lemma 7.118. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Let $i \in [n]$. Then,

$$\ell_i(\sigma) = \sum_{j=i+1}^{n} [\sigma(i) > \sigma(j)].$$

Proof of Lemma 7.118. Recall that $\ell_i(\sigma)$ is the number of all $j \in \{i+1, i+2, \ldots, n\}$ such that $\sigma(i) > \sigma(j)$ (by the definition of $\ell_i(\sigma)$). In other words,

$$\ell_i(\sigma) = |\{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}|.$$

Comparing this with

$$\begin{aligned}
&\underbrace{\sum_{j=i+1}^{n}}_{=\sum_{j\in\{i+1,i+2,\ldots,n\}}} [\sigma(i) > \sigma(j)] \\
&= \sum_{j\in\{i+1,i+2,\ldots,n\}} [\sigma(i) > \sigma(j)] \\
&= \sum_{\substack{j\in\{i+1,i+2,\ldots,n\};\\ \sigma(i)>\sigma(j)}} \underbrace{[\sigma(i) > \sigma(j)]}_{\substack{=1\\ (\text{since } \sigma(i)>\sigma(j))}} + \sum_{\substack{j\in\{i+1,i+2,\ldots,n\};\\ \text{not } \sigma(i)>\sigma(j)}} \underbrace{[\sigma(i) > \sigma(j)]}_{\substack{=0\\ (\text{since we don't have } \sigma(i)>\sigma(j))}} \\
&\qquad \left(\begin{array}{c}\text{since each } j \in \{i+1,i+2,\ldots,n\} \text{ satisfies either } (\sigma(i) > \sigma(j)) \\ \text{or } (\text{not } \sigma(i) > \sigma(j)) \text{ (but not both)}\end{array}\right) \\
&= \sum_{\substack{j\in\{i+1,i+2,\ldots,n\};\\ \sigma(i)>\sigma(j)}} 1 + \underbrace{\sum_{\substack{j\in\{i+1,i+2,\ldots,n\};\\ \text{not } \sigma(i)>\sigma(j)}} 0}_{=0} = \sum_{\substack{j\in\{i+1,i+2,\ldots,n\};\\ \sigma(i)>\sigma(j)}} 1 \\
&= |\{j \in \{i+1,i+2,\ldots,n\} \mid \sigma(i) > \sigma(j)\}| \cdot 1 \\
&= |\{j \in \{i+1,i+2,\ldots,n\} \mid \sigma(i) > \sigma(j)\}|,
\end{aligned}$$

we obtain $\ell_i(\sigma) = \sum_{j=i+1}^{n} [\sigma(i) > \sigma(j)]$. This proves Lemma 7.118. □

Lemma 7.119. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Let $u \in [n]$ and $v \in [n]$ be such that $u < v$. Let $i \in \{u+1, u+2, \ldots, v-1\}$. Then,

$$\begin{aligned}
&[\sigma(u) > \sigma(i)] - [\sigma(v) > \sigma(i)] + [\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)] \\
&= 2[\sigma(u) > \sigma(i) > \sigma(v)] - 2[\sigma(v) > \sigma(i) > \sigma(u)].
\end{aligned}$$

Proof of Lemma 7.119. We know that σ is a permutation (since $\sigma \in S_n$), thus a bijective map, thus an injective map. From $i \in \{u+1, u+2, \ldots, v-1\}$, we obtain $u+1 \leq i \leq v-1$. From $u < u+1 \leq i$, we obtain $u \neq i$ and thus $\sigma(u) \neq \sigma(i)$ (since σ is injective). From $i \leq v-1 < v$, we obtain $i \neq v$ and thus $\sigma(i) \neq \sigma(v)$ (since σ is injective).

From $\sigma(u) \neq \sigma(i)$, we conclude that either $\sigma(u) < \sigma(i)$ or $\sigma(u) > \sigma(i)$. From $\sigma(i) \neq \sigma(v)$, we conclude that either $\sigma(i) < \sigma(v)$ or $\sigma(i) > \sigma(v)$. Combining the previous two sentences, we conclude that we must be in one of the following four cases:

Case 1: We have $\sigma(u) < \sigma(i)$ and $\sigma(i) < \sigma(v)$.
Case 2: We have $\sigma(u) < \sigma(i)$ and $\sigma(i) > \sigma(v)$.
Case 3: We have $\sigma(u) > \sigma(i)$ and $\sigma(i) < \sigma(v)$.
Case 4: We have $\sigma(u) > \sigma(i)$ and $\sigma(i) > \sigma(v)$.

We can now verify the claim of Lemma 7.119 in each of these four Cases by inspection:

- In Case 1, the equality claimed in Lemma 7.119 boils down to $0 - 1 + 0 - 1 = 2 \cdot 0 - 2 \cdot 1$, which is correct.
- In Case 2, the equality claimed in Lemma 7.119 boils down to $0 - 0 + 1 - 1 = 2 \cdot 0 - 2 \cdot 0$, which is correct.
- In Case 3, the equality claimed in Lemma 7.119 boils down to $1 - 1 + 0 - 0 = 2 \cdot 0 - 2 \cdot 0$, which is correct.
- In Case 4, the equality claimed in Lemma 7.119 boils down to $1 - 0 + 1 - 0 = 2 \cdot 1 - 2 \cdot 0$, which is correct.

Thus, the equality claimed in Lemma 7.119 holds in all four Cases 1, 2, 3 and 4. Hence, Lemma 7.119 is proven. □

Lemma 7.120. Let $n \in \mathbb{N}$. Let $u \in [n]$ and $v \in [n]$ be such that $u < v$. Let $\sigma \in S_n$ and $\tau \in S_n$ be such that $\tau = \sigma \circ t_{u,v}$. Then:

(a) We have $\ell_i(\sigma) = \ell_i(\tau)$ for each $i \in \{1, 2, \ldots, u-1\}$.

(b) We have

$$\ell_u(\sigma) = \ell_v(\tau) + \sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] + [\sigma(u) > \sigma(v)].$$

(c) We have

$$\ell_i(\sigma) = \ell_i(\tau) + [\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)]$$

for each $i \in \{u+1, u+2, \ldots, v-1\}$.

(d) We have

$$\ell_v(\sigma) = \ell_u(\tau) - \sum_{i=u+1}^{v-1} [\sigma(v) > \sigma(i)] - [\sigma(v) > \sigma(u)].$$

(e) We have $\ell_i(\sigma) = \ell_i(\tau)$ for each $i \in \{v+1, v+2, \ldots, n\}$.

(f) We have

$$\ell(\sigma) = \ell(\tau) + [\sigma(u) > \sigma(v)] - [\sigma(v) > \sigma(u)] + \sum_{i=u+1}^{v-1} (2[\sigma(u) > \sigma(i) > \sigma(v)] - 2[\sigma(v) > \sigma(i) > \sigma(u)]).$$

Proof of Lemma 7.120. We have $u \in [n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$) and $v \in \{1, 2, \ldots, n\}$ (similarly). From $u \in \{1, 2, \ldots, n\}$, we obtain $u \geq 1$. From $v \in$

$\{1,2,\ldots,n\}$, we obtain $v \leq n$. Thus, $1 \leq u < v \leq n$. Note that u and v are distinct (since $u < v$). Also, from $u < v$, we obtain $u \leq v-1$ (since u and v are integers), and thus $u+1 \leq v$.

Recall that $t_{u,v}$ is the permutation in S_n which switches u with v while leaving all other elements of $\{1,2,\ldots,n\}$ unchanged (by the definition of $t_{u,v}$). In other words, $t_{u,v}$ is the permutation in S_n that satisfies $t_{u,v}(u) = v$ and $t_{u,v}(v) = u$ and

$$\left(t_{u,v}(i) = i \qquad \text{for each } i \in \{1,2,\ldots,n\} \setminus \{u,v\}\right). \tag{866}$$

Now,

$$\underbrace{\tau}_{=\sigma\circ t_{u,v}}(u) = (\sigma \circ t_{u,v})(u) = \sigma\left(\underbrace{t_{u,v}(u)}_{=v}\right) = \sigma(v) \qquad \text{and}$$

$$\underbrace{\tau}_{=\sigma\circ t_{u,v}}(v) = (\sigma \circ t_{u,v})(v) = \sigma\left(\underbrace{t_{u,v}(v)}_{=u}\right) = \sigma(u).$$

Moreover, each $i \in \{1,2,\ldots,n\} \setminus \{u,v\}$ satisfies

$$\underbrace{\tau}_{=\sigma\circ t_{u,v}}(i) = (\sigma \circ t_{u,v})(i) = \sigma\left(\underbrace{t_{u,v}(i)}_{\substack{=i \\ (\text{by (866)})}}\right) = \sigma(i). \tag{867}$$

From this, we obtain the following consequences:

- Each $i \in \{1,2,\ldots,u-1\}$ satisfies

$$\tau(i) = \sigma(i). \tag{868}$$

  [*Proof of (868):* Let $i \in \{1,2,\ldots,u-1\}$. Thus, $i \leq u-1 < u$, so that $i \neq u$. Also, $i < u < v$, so that $i \neq v$. Also, $i \in \{1,2,\ldots,u-1\} \subseteq \{1,2,\ldots,n\}$ (since $u-1 \leq u \leq n$). Combining $i \neq u$ with $i \neq v$, we conclude that $i \notin \{u,v\}$. Thus, $i \in \{1,2,\ldots,n\} \setminus \{u,v\}$ (since $i \in \{1,2,\ldots,n\}$). Hence, (867) yields $\tau(i) = \sigma(i)$. This proves (868).]

- Each $i \in \{u+1,u+2,\ldots,v-1\}$ satisfies

$$\tau(i) = \sigma(i). \tag{869}$$

  [*Proof of (869):* Let $i \in \{u+1,u+2,\ldots,v-1\}$. Thus, $u+1 \leq i \leq v-1$. We have $u+1 \leq i$, thus $i \geq u+1 > u$ and therefore $i \neq u$. Also, $i \leq v-1 < v$ and thus $i \neq v$. Moreover, $i > u \geq 1$ and thus $i \geq 1$. Also, $i < v \leq n$ and thus $i \leq n$. Thus, $i \in \{1,2,\ldots,n\}$ (since $i \geq 1$). Combining $i \neq u$ with $i \neq v$, we conclude that $i \notin \{u,v\}$. Thus, $i \in \{1,2,\ldots,n\} \setminus \{u,v\}$ (since $i \in \{1,2,\ldots,n\}$). Hence, (867) yields $\tau(i) = \sigma(i)$. This proves (869).]

- Each $j \in \{v+1, v+2, \ldots, n\}$ satisfies
$$\tau(j) = \sigma(j). \tag{870}$$
[*Proof of (870):* Let $j \in \{v+1, v+2, \ldots, n\}$. Thus, $j \geq v+1 > v$, so that $j \neq v$. Also, $j > v > u$ (since $u < v$), so that $j \neq u$. Also, $j \in \{v+1, v+2, \ldots, n\} \subseteq \{1, 2, \ldots, n\}$ (since $v+1 \geq v \geq 1$). Combining $j \neq u$ with $j \neq v$, we conclude that $j \notin \{u, v\}$. Thus, $j \in \{1, 2, \ldots, n\} \setminus \{u, v\}$ (since $j \in \{1, 2, \ldots, n\}$). Hence, (867) (applied to $i = j$) yields $\tau(j) = \sigma(j)$. This proves (870).]

(a) Let $i \in \{1, 2, \ldots, u-1\}$. Thus, (868) yields $\tau(i) = \sigma(i)$.
Also, each $j \in [i]$ yields
$$\tau(j) = \sigma(j). \tag{871}$$
[*Proof of (871):* Let $j \in [i]$. Thus, $j \in [i] = \{1, 2, \ldots, i\}$ (by the definition of $[i]$). Hence, $j \geq 1$ and $j \leq i \leq u-1$ (since $i \in \{1, 2, \ldots, u-1\}$), so that $j \in \{1, 2, \ldots, u-1\}$ (because $j \geq 1$). Thus, (868) (applied to j instead of i) yields $\tau(j) = \sigma(j)$. This proves (871).]

Now,
$$\tau([i]) = \left\{ \underbrace{\tau(j)}_{\substack{=\sigma(j) \\ \text{(by (871))}}} \mid j \in [i] \right\} = \{\sigma(j) \mid j \in [i]\} = \sigma([i]).$$
Lemma 5.48 **(a)** yields $\ell_i(\sigma) = |[\sigma(i) - 1] \setminus \sigma([i])|$. The same argument (applied to τ instead of σ) yields $\ell_i(\tau) = |[\tau(i) - 1] \setminus \tau([i])|$. Hence,
$$\ell_i(\tau) = \left| \left[\underbrace{\tau(i)}_{=\sigma(i)} - 1 \right] \setminus \underbrace{\tau([i])}_{=\sigma([i])} \right| = |[\sigma(i) - 1] \setminus \sigma([i])| = \ell_i(\sigma)$$
(since $\ell_i(\sigma) = |[\sigma(i) - 1] \setminus \sigma([i])|$). In other words, $\ell_i(\sigma) = \ell_i(\tau)$. This proves Lemma 7.120 **(a)**.

(b) Lemma 7.118 (applied to $i = u$) yields
$$\ell_u(\sigma) = \sum_{j=u+1}^{n} [\sigma(u) > \sigma(j)] = \sum_{j=u+1}^{v-1} [\sigma(u) > \sigma(j)] + \underbrace{\sum_{j=v}^{n} [\sigma(u) > \sigma(j)]}_{\substack{=[\sigma(u)>\sigma(v)] + \sum_{j=v+1}^{n} [\sigma(u)>\sigma(j)] \\ \text{(here, we have split off the addend for } j=v \\ \text{from the sum, since } v \leq n)}}$$
$$\left(\begin{array}{c} \text{here, we have split the sum at } j = v, \\ \text{because } u+1 \leq v \leq n \end{array} \right)$$
$$= \sum_{j=u+1}^{v-1} [\sigma(u) > \sigma(j)] + [\sigma(u) > \sigma(v)] + \sum_{j=v+1}^{n} [\sigma(u) > \sigma(j)]. \tag{872}$$

But Lemma 7.118 (applied to v and τ instead of i and σ) yields

$$\ell_v(\tau) = \sum_{j=v+1}^{n} \left[\underbrace{\tau(v)}_{=\sigma(u)} > \underbrace{\tau(j)}_{\substack{=\sigma(j) \\ \text{(by (870))}}} \right] = \sum_{j=v+1}^{n} [\sigma(u) > \sigma(j)]. \qquad (873)$$

Hence, (872) becomes

$$\begin{aligned} \ell_u(\sigma) &= \underbrace{\sum_{j=u+1}^{v-1} [\sigma(u) > \sigma(j)]}_{\substack{= \sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] \\ \text{(here, we have renamed} \\ \text{the summation index } j \text{ as } i)}} + [\sigma(u) > \sigma(v)] + \underbrace{\sum_{j=v+1}^{n} [\sigma(u) > \sigma(j)]}_{\substack{=\ell_v(\tau) \\ \text{(by (873))}}} \\ &= \sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] + [\sigma(u) > \sigma(v)] + \ell_v(\tau) \\ &= \ell_v(\tau) + \sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] + [\sigma(u) > \sigma(v)]. \end{aligned}$$

This proves Lemma 7.120 **(b)**.

(c) Let $i \in \{u+1, u+2, \ldots, v-1\}$. Thus, $u+1 \leq i \leq v-1$. Hence, $v \in \{i+1, i+2, \ldots, n\}$ [424].

Furthermore, each $j \in \{i+1, i+2, \ldots, n\}$ satisfying $j \neq v$ must satisfy

$$\tau(j) = \sigma(j). \qquad (874)$$

[*Proof of (874):* Let $j \in \{i+1, i+2, \ldots, n\}$ be such that $j \neq v$.

From $j \in \{i+1, i+2, \ldots, n\}$, we obtain $j \geq i+1 > i > u$ (since $u < u+1 \leq i$) and therefore $j \neq u$. Also, $j \in \{i+1, i+2, \ldots, n\} \subseteq \{1, 2, \ldots, n\}$ (since $i+1 \geq i > u \geq 1$). Combining $j \neq u$ with $j \neq v$, we conclude that $j \notin \{u, v\}$. Thus, $j \in \{1, 2, \ldots, n\} \setminus \{u, v\}$ (since $j \in \{1, 2, \ldots, n\}$). Hence, (867) (applied to j instead of i) yields $\tau(j) = \sigma(j)$. This proves (874).]

Lemma 7.118 yields

$$\begin{aligned} \ell_i(\sigma) &= \underbrace{\sum_{j=i+1}^{n}}_{= \sum_{j \in \{i+1, i+2, \ldots, n\}}} [\sigma(i) > \sigma(j)] = \sum_{j \in \{i+1, i+2, \ldots, n\}} [\sigma(i) > \sigma(j)] \\ &= [\sigma(i) > \sigma(v)] + \sum_{\substack{j \in \{i+1, i+2, \ldots, n\}; \\ j \neq v}} [\sigma(i) > \sigma(j)] \qquad (875) \end{aligned}$$

[424] *Proof.* We have $i \leq v-1$, so that $v \geq i+1$. Combining this with $v \leq n$, we obtain $v \in \{i+1, i+2, \ldots, n\}$ (since v is an integer).

(here, we have split off the addend for $j = v$ from the sum, since we have $v \in \{i+1, i+2, \ldots, n\}$). The same argument (but applied to τ instead of σ) yields

$$\ell_i(\tau) = [\tau(i) > \tau(v)] + \sum_{\substack{j \in \{i+1,i+2,\ldots,n\};\\ j \neq v}} [\tau(i) > \tau(j)].$$

Thus,

$$\begin{aligned}
\ell_i(\tau) &= \left[\underbrace{\tau(i)}_{\substack{=\sigma(i)\\ \text{(by (869))}}} > \underbrace{\tau(v)}_{=\sigma(u)}\right] + \sum_{\substack{j \in \{i+1,i+2,\ldots,n\};\\ j \neq v}} \left[\underbrace{\tau(i)}_{\substack{=\sigma(i)\\ \text{(by (869))}}} > \underbrace{\tau(j)}_{\substack{=\sigma(j)\\ \text{(by (874))}}}\right] \\
&= [\sigma(i) > \sigma(u)] + \sum_{\substack{j \in \{i+1,i+2,\ldots,n\};\\ j \neq v}} [\sigma(i) > \sigma(j)].
\end{aligned}$$

Subtracting this equality from (875), we obtain

$$\begin{aligned}
\ell_i(\sigma) - \ell_i(\tau) &= \left([\sigma(i) > \sigma(v)] + \sum_{\substack{j \in \{i+1,i+2,\ldots,n\};\\ j \neq v}} [\sigma(i) > \sigma(j)]\right) \\
&\qquad - \left([\sigma(i) > \sigma(u)] + \sum_{\substack{j \in \{i+1,i+2,\ldots,n\};\\ j \neq v}} [\sigma(i) > \sigma(j)]\right) \\
&= [\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)].
\end{aligned}$$

Adding $\ell_i(\tau)$ to both sides of this equality, we find

$$\ell_i(\sigma) = \ell_i(\tau) + [\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)].$$

This proves Lemma 7.120 **(c)**.

(d) Lemma 7.75 **(d)** (applied to u and v instead of i and j) yields $t_{u,v} \circ t_{u,v} = \mathrm{id}$. Now,

$$\underbrace{\tau}_{=\sigma \circ t_{u,v}} \circ t_{u,v} = \sigma \circ \underbrace{t_{u,v} \circ t_{u,v}}_{=\mathrm{id}} = \sigma,$$

so that $\sigma = \tau \circ t_{u,v}$. Hence, we can apply Lemma 7.120 **(b)** to σ and τ instead of τ and σ. We thus find

$$\begin{aligned}
\ell_u(\tau) &= \ell_v(\sigma) + \sum_{i=u+1}^{v-1} \left[\underbrace{\tau(u)}_{=\sigma(v)} > \underbrace{\tau(i)}_{\substack{=\sigma(i)\\ \text{(by (869))}}}\right] + \left[\underbrace{\tau(u)}_{=\sigma(v)} > \underbrace{\tau(v)}_{=\sigma(u)}\right] \\
&= \ell_v(\sigma) + \sum_{i=u+1}^{v-1} [\sigma(v) > \sigma(i)] + [\sigma(v) > \sigma(u)].
\end{aligned}$$

Solving this equation for $\ell_v(\sigma)$, we obtain

$$\ell_v(\sigma) = \ell_u(\tau) - \sum_{i=u+1}^{v-1} [\sigma(v) > \sigma(i)] - [\sigma(v) > \sigma(u)].$$

This proves Lemma 7.120 **(d)**.

(e) Let $i \in \{v+1, v+2, \ldots, n\}$. Then, (870) (applied to $j = i$) yields $\tau(i) = \sigma(i)$. Also, each $j \in \{i+1, i+2, \ldots, n\}$ satisfies

$$\tau(j) = \sigma(j). \tag{876}$$

[*Proof of (876):* Let $j \in \{i+1, i+2, \ldots, n\}$. Thus, $j \geq i+1 > i \geq v+1$ (since $i \in \{v+1, v+2, \ldots, n\}$), so that $j \geq v+1$. Also, $j \leq n$ (since $j \in \{i+1, i+2, \ldots, n\}$). Combining $j \geq v+1$ with $j \leq n$, we obtain $j \in \{v+1, v+2, \ldots, n\}$. Thus, (870) yields $\tau(j) = \sigma(j)$. This proves (876).]

Lemma 7.118 yields

$$\ell_i(\sigma) = \sum_{j=i+1}^{n} [\sigma(i) > \sigma(j)].$$

Lemma 7.118 (applied to τ instead of σ) yields

$$\ell_i(\tau) = \sum_{j=i+1}^{n} \left[\underbrace{\tau(i)}_{=\sigma(i)} > \underbrace{\tau(j)}_{\substack{=\sigma(j) \\ \text{(by (876))}}} \right] = \sum_{j=i+1}^{n} [\sigma(i) > \sigma(j)].$$

Comparing these two equalities, we obtain $\ell_i(\sigma) = \ell_i(\tau)$. This proves Lemma 7.120 **(e)**.

(f) Proposition 5.46 yields

$$\ell(\sigma) = \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma) = \sum_{i=1}^{n} \ell_i(\sigma)$$
$$= \underbrace{\sum_{i=1}^{v-1} \ell_i(\sigma)}_{\substack{=\sum_{i=1}^{u-1}\ell_i(\sigma)+\sum_{i=u}^{v-1}\ell_i(\sigma) \\ \text{(here, we have split the sum at } i=u\text{,} \\ \text{since } 1\leq u\leq v-1\text{)}}} + \underbrace{\sum_{i=v}^{n} \ell_i(\sigma)}_{\substack{=\ell_v(\sigma)+\sum_{i=v+1}^{n}\ell_i(\sigma) \\ \text{(here, we have split off the addend for } i=v \text{ from the sum,} \\ \text{since } v\leq n\text{)}}}$$
$$\left(\text{here, we have split the sum at } i = v\text{, because } 1 \leq v \leq n\right)$$
$$= \sum_{i=1}^{u-1} \ell_i(\sigma) + \underbrace{\sum_{i=u}^{v-1} \ell_i(\sigma)}_{\substack{=\ell_u(\sigma)+\sum_{i=u+1}^{v-1}\ell_i(\sigma) \\ \text{(here, we have split off the addend for } i=u \text{ from the sum,} \\ \text{since } u\leq v-1\text{)}}} + \ell_v(\sigma) + \sum_{i=v+1}^{n} \ell_i(\sigma)$$
$$= \sum_{i=1}^{u-1} \ell_i(\sigma) + \ell_u(\sigma) + \sum_{i=u+1}^{v-1} \ell_i(\sigma) + \ell_v(\sigma) + \sum_{i=v+1}^{n} \ell_i(\sigma). \tag{877}$$

The same argument (applied to τ instead of σ) yields

$$\ell(\tau) = \sum_{i=1}^{u-1} \ell_i(\tau) + \ell_u(\tau) + \sum_{i=u+1}^{v-1} \ell_i(\tau) + \ell_v(\tau) + \sum_{i=v+1}^{n} \ell_i(\tau). \tag{878}$$

But (877) becomes

$$\begin{aligned}
\ell(\sigma) &= \sum_{i=1}^{u-1} \underbrace{\ell_i(\sigma)}_{\substack{=\ell_i(\tau) \\ \text{(by Lemma 7.120 (a))}}} + \underbrace{\ell_u(\sigma)}_{\substack{=\ell_v(\tau)+\sum_{i=u+1}^{v-1}[\sigma(u)>\sigma(i)]+[\sigma(u)>\sigma(v)] \\ \text{(by Lemma 7.120 (b))}}} \\
&\qquad + \sum_{i=u+1}^{v-1} \underbrace{\ell_i(\sigma)}_{\substack{=\ell_i(\tau)+[\sigma(i)>\sigma(v)]-[\sigma(i)>\sigma(u)] \\ \text{(by Lemma 7.120 (c))}}} \\
&\qquad + \underbrace{\ell_v(\sigma)}_{\substack{=\ell_u(\tau)-\sum_{i=u+1}^{v-1}[\sigma(v)>\sigma(i)]-[\sigma(v)>\sigma(u)] \\ \text{(by Lemma 7.120 (d))}}} + \sum_{i=v+1}^{n} \underbrace{\ell_i(\sigma)}_{\substack{=\ell_i(\tau) \\ \text{(by Lemma 7.120 (e))}}} \\
&= \sum_{i=1}^{u-1} \ell_i(\tau) + \ell_v(\tau) + \sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] + [\sigma(u) > \sigma(v)] \\
&\qquad + \underbrace{\sum_{i=u+1}^{v-1} (\ell_i(\tau) + [\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)])}_{=\sum_{i=u+1}^{v-1} \ell_i(\tau) + \sum_{i=u+1}^{v-1} ([\sigma(i)>\sigma(v)]-[\sigma(i)>\sigma(u)])} \\
&\qquad + \ell_u(\tau) - \sum_{i=u+1}^{v-1} [\sigma(v) > \sigma(i)] - [\sigma(v) > \sigma(u)] + \sum_{i=v+1}^{n} \ell_i(\tau) \\
&= \sum_{i=1}^{u-1} \ell_i(\tau) + \ell_v(\tau) + \sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] + [\sigma(u) > \sigma(v)] \\
&\qquad + \sum_{i=u+1}^{v-1} \ell_i(\tau) + \sum_{i=u+1}^{v-1} ([\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)]) \\
&\qquad + \ell_u(\tau) - \sum_{i=u+1}^{v-1} [\sigma(v) > \sigma(i)] - [\sigma(v) > \sigma(u)] + \sum_{i=v+1}^{n} \ell_i(\tau)
\end{aligned}$$

$$= \underbrace{\sum_{i=1}^{u-1} \ell_i(\tau) + \ell_u(\tau) + \sum_{i=u+1}^{v-1} \ell_i(\tau) + \ell_v(\tau) + \sum_{i=v+1}^{n} \ell_i(\tau)}_{\substack{=\ell(\tau) \\ \text{(by (878))}}}$$
$$+ [\sigma(u) > \sigma(v)] - [\sigma(v) > \sigma(u)]$$
$$+ \underbrace{\sum_{i=u+1}^{v-1} [\sigma(u) > \sigma(i)] + \sum_{i=u+1}^{v-1} ([\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)]) - \sum_{i=u+1}^{v-1} [\sigma(v) > \sigma(i)]}_{=\sum_{i=u+1}^{v-1} ([\sigma(u)>\sigma(i)]+([\sigma(i)>\sigma(v)]-[\sigma(i)>\sigma(u)])-[\sigma(v)>\sigma(i)])}$$
$$= \ell(\tau) + [\sigma(u) > \sigma(v)] - [\sigma(v) > \sigma(u)]$$
$$+ \sum_{i=u+1}^{v-1} \underbrace{([\sigma(u) > \sigma(i)] + ([\sigma(i) > \sigma(v)] - [\sigma(i) > \sigma(u)]) - [\sigma(v) > \sigma(i)])}_{\substack{=[\sigma(u)>\sigma(i)]-[\sigma(v)>\sigma(i)]+[\sigma(i)>\sigma(v)]-[\sigma(i)>\sigma(u)] \\ =2[\sigma(u)>\sigma(i)>\sigma(v)]-2[\sigma(v)>\sigma(i)>\sigma(u)] \\ \text{(by Lemma 7.119)}}}$$
$$= \ell(\tau) + [\sigma(u) > \sigma(v)] - [\sigma(v) > \sigma(u)]$$
$$+ \sum_{i=u+1}^{v-1} (2[\sigma(u) > \sigma(i) > \sigma(v)] - 2[\sigma(v) > \sigma(i) > \sigma(u)]).$$

This proves Lemma 7.120 **(f)**. □

Our next lemma is precisely the claim of Exercise 5.20, with i, j and k renamed as u, v and i:

Lemma 7.121. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Let u and v be two elements of $[n]$ such that $u < v$ and $\sigma(u) > \sigma(v)$. Let Q be the set of all $i \in \{u+1, u+2, \ldots, v-1\}$ satisfying $\sigma(u) > \sigma(i) > \sigma(v)$. Then,

$$\ell(\sigma \circ t_{u,v}) = \ell(\sigma) - 2|Q| - 1.$$

Proof of Lemma 7.121. Recall that Q is the set of all $i \in \{u+1, u+2, \ldots, v-1\}$ satisfying $\sigma(u) > \sigma(i) > \sigma(v)$. In other words,

$$Q = \{i \in \{u+1, u+2, \ldots, v-1\} \mid \sigma(u) > \sigma(i) > \sigma(v)\}. \qquad (879)$$

Define a permutation $\tau \in S_n$ by $\tau = \sigma \circ t_{u,v}$. We have $\sigma(u) > \sigma(v)$; thus, we don't have $\sigma(v) > \sigma(u)$. Hence, $[\sigma(v) > \sigma(u)] = 0$.

For any $i \in \{u+1, u+2, \ldots, v-1\}$, we have

$$[\sigma(v) > \sigma(i) > \sigma(u)] = 0 \qquad (880)$$

(since we don't have $\sigma(v) > \sigma(i) > \sigma(u)$ (because we don't have $\sigma(v) > \sigma(u)$)).

Now, Lemma 7.120 **(f)** yields

$$\begin{aligned}
\ell(\sigma) &= \ell(\tau) + \underbrace{[\sigma(u) > \sigma(v)]}_{\substack{=1 \\ (\text{since } \sigma(u)>\sigma(v))}} - \underbrace{[\sigma(v) > \sigma(u)]}_{=0} \\
&\qquad + \underbrace{\sum_{i=u+1}^{v-1}}_{= \sum_{i\in\{u+1,u+2,\ldots,v-1\}}} \left(2[\sigma(u) > \sigma(i) > \sigma(v)] - 2\underbrace{[\sigma(v) > \sigma(i) > \sigma(u)]}_{\substack{=0 \\ (\text{by } (880))}} \right) \\
&= \ell(\tau) + \underbrace{1-0}_{=1} + \sum_{i\in\{u+1,u+2,\ldots,v-1\}} \underbrace{(2[\sigma(u) > \sigma(i) > \sigma(v)] - 2\cdot 0)}_{=2[\sigma(u)>\sigma(i)>\sigma(v)]} \\
&= \ell(\tau) + 1 + \underbrace{\sum_{i\in\{u+1,u+2,\ldots,v-1\}} 2[\sigma(u) > \sigma(i) > \sigma(v)]}_{=2\sum_{i\in\{u+1,u+2,\ldots,v-1\}}[\sigma(u)>\sigma(i)>\sigma(v)]} \\
&= \ell(\tau) + 1 + 2\sum_{i\in\{u+1,u+2,\ldots,v-1\}} [\sigma(u) > \sigma(i) > \sigma(v)].
\end{aligned}$$

In view of

$$\begin{aligned}
&\sum_{i\in\{u+1,u+2,\ldots,v-1\}} [\sigma(u) > \sigma(i) > \sigma(v)] \\
&= \sum_{\substack{i\in\{u+1,u+2,\ldots,v-1\}; \\ \sigma(u)>\sigma(i)>\sigma(v)}} \underbrace{[\sigma(u) > \sigma(i) > \sigma(v)]}_{\substack{=1 \\ (\text{since } \sigma(u)>\sigma(i)>\sigma(v))}} + \sum_{\substack{i\in\{u+1,u+2,\ldots,v-1\}; \\ \text{not } \sigma(u)>\sigma(i)>\sigma(v)}} \underbrace{[\sigma(u) > \sigma(i) > \sigma(v)]}_{\substack{=0 \\ (\text{since we don't have } \sigma(u)>\sigma(i)>\sigma(v))}} \\
&\qquad \left(\begin{array}{c} \text{since each } i \in \{u+1, u+2, \ldots, v-1\} \text{ satisfies either } (\sigma(u) > \sigma(i) > \sigma(v)) \\ \text{or } (\text{not } \sigma(u) > \sigma(i) > \sigma(v)) \text{ (but not both)} \end{array} \right) \\
&= \sum_{\substack{i\in\{u+1,u+2,\ldots,v-1\}; \\ \sigma(u)>\sigma(i)>\sigma(v)}} 1 + \underbrace{\sum_{\substack{i\in\{u+1,u+2,\ldots,v-1\}; \\ \text{not } \sigma(u)>\sigma(i)>\sigma(v)}} 0}_{=0} = \sum_{\substack{i\in\{u+1,u+2,\ldots,v-1\}; \\ \sigma(u)>\sigma(i)>\sigma(v)}} 1 \\
&= \left| \underbrace{\{i \in \{u+1, u+2, \ldots, v-1\} \mid \sigma(u) > \sigma(i) > \sigma(v)\}}_{\substack{=Q \\ (\text{by } (879))}} \right| \cdot 1 = |Q| \cdot 1 = |Q|,
\end{aligned}$$

this becomes

$$\ell(\sigma) = \ell(\tau) + 1 + 2\underbrace{\sum_{i\in\{u+1,u+2,\ldots,v-1\}} [\sigma(u) > \sigma(i) > \sigma(v)]}_{=|Q|} = \ell(\tau) + 1 + 2|Q|.$$

Solving this equation for $\ell(\tau)$, we find

$$\ell(\tau) = \ell(\sigma) - 1 - 2|Q| = \ell(\sigma) - 2|Q| - 1.$$

In view of $\tau = \sigma \circ t_{u,v}$, this rewrites as $\ell(\sigma \circ t_{u,v}) = \ell(\sigma) - 2|Q| - 1$. Thus, Lemma 7.121 is proven. □

Solution to Exercise 5.20. We obtain the statement of Exercise 5.20 if we rename the variables u, v and i as i, j and k in Lemma 7.121. Thus, Exercise 5.20 is solved. □

7.60. Solution to Exercise 5.21

Throughout Section 7.60, we shall use the same notations that were used in Section 5.8. We shall also use Definition 7.104.

We begin with a simple lemma that will help us resolve parts **(a)** and **(b)** of Exercise 5.21:

Lemma 7.122. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Let $i \in [n]$.

(a) The permutation $c_{i,i+\ell_i(\sigma)} \in S_n$ is well-defined.

(b) We have $c_{i,i+\ell_i(\sigma)} = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$, where $i' = i + \ell_i(\sigma)$.

Proof of Lemma 7.122. We know (from Corollary 7.105 **(a)**) that the permutation $c_{u,v}$ is well-defined whenever u and v are two elements of $[n]$ such that $u \leq v$.

Now, $i \in [n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$). Also, Proposition 5.47 yields

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \in H = [n-1]_0 \times [n-2]_0 \times \cdots \times [n-n]_0$$

(by the definition of H). In other words, $\ell_j(\sigma) \in [n-j]_0$ for each $j \in \{1, 2, \ldots, n\}$. Applying this to $j = i$, we obtain

$$\ell_i(\sigma) \in [n-i]_0 = \{0, 1, \ldots, n-i\}$$

(by the definition of $[n-i]_0$). Hence, $\ell_i(\sigma) \geq 0$ and $\ell_i(\sigma) \leq n - i$. Now, $i + \underbrace{\ell_i(\sigma)}_{\geq 0} \geq i$, so that $i \leq i + \ell_i(\sigma)$.

Also, $i + \underbrace{\ell_i(\sigma)}_{\geq 0} \geq i \geq 1$ (since $i \in \{1, 2, \ldots, n\}$). Combining this with $i + \underbrace{\ell_i(\sigma)}_{\leq n-i} \leq i + n - i = n$, we obtain $i + \ell_i(\sigma) \in \{1, 2, \ldots, n\} = [n]$. Hence, i and $i + \ell_i(\sigma)$ are two elements of $[n]$ such that $i \leq i + \ell_i(\sigma)$. Hence, Lemma 7.105 **(a)** (applied to $u = i$ and $v = i + \ell_i(\sigma)$) yields that the permutation $c_{i,i+\ell_i(\sigma)}$ is well-defined. This proves Lemma 7.122 **(a)**.

(b) Let $i' = i + \ell_i(\sigma)$. We must prove that $c_{i,i+\ell_i(\sigma)} = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$.

We have $i' = i + \ell_i(\sigma) \geq i$, so that $i \leq i'$. Also, $i \in [n]$ and $i' = i + \ell_i(\sigma) \in [n]$. Hence, Lemma 7.105 **(b)** (applied to $u = i$ and $v = i'$) yields $c_{i,i'} = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$. But $i + \ell_i(\sigma) = i'$ (since $i' = i + \ell_i(\sigma)$); thus, $c_{i,i+\ell_i(\sigma)} = c_{i,i'} = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$. This proves Lemma 7.122 **(b)**. □

Next, we state a property of the permutations $c_{u,v}$ that is essentially clear from their definition:

Lemma 7.123. Let $n \in \mathbb{N}$. Let u and v be two elements of $[n]$ such that $u \leq v$. Consider the permutation $c_{u,v} \in S_n$ defined in Definition 7.104. Then:

(a) We have $c_{u,v}(q) = q$ for each $q \in [n] \setminus \{u, u+1, \ldots, v\}$.

(b) We have $c_{u,v}(q) = q - 1$ for each $q \in \{u+1, u+2, \ldots, v\}$.

(c) We have $c_{u,v}(u) = v$.

Proof of Lemma 7.123. The definition of $c_{u,v}$ yields $c_{u,v} = \operatorname{cyc}_{v,v-1,v-2,\ldots,u}$. But the $(v-u+1)$-cycle $\operatorname{cyc}_{v,v-1,v-2,\ldots,u}$ was defined as the permutation in S_n which sends $v, v-1, \ldots, u+1, u$ to $v-1, v-2, \ldots, u, v$, respectively, while leaving all other elements of $[n]$ fixed. In other words, $c_{u,v}$ is the permutation in S_n which sends $v, v-1, \ldots, u+1, u$ to $v-1, v-2, \ldots, u, v$, respectively, while leaving all other elements of $[n]$ fixed. This immediately yields all three parts of Lemma 7.123. $\square$

Corollary 7.124. Let $n \in \mathbb{N}$. Let u and v be two elements of $[n]$ such that $u \leq v$. Consider the permutation $c_{u,v} \in S_n$ defined in Definition 7.104. Let $k \in [n]$ be such that $k \neq u$. Then, $c_{u,v}(k) \in \{k-1, k\}$.

Proof of Corollary 7.124. We are in one of the following two cases:

Case 1: We have $k \in \{u, u+1, \ldots, v\}$.

Case 2: We don't have $k \in \{u, u+1, \ldots, v\}$.

Let us first consider Case 1. In this case, we have $k \in \{u, u+1, \ldots, v\}$. Combining this with $k \neq u$, we obtain $k \in \{u, u+1, \ldots, v\} \setminus \{u\} = \{u+1, u+2, \ldots, v\}$. Hence, Lemma 7.123 **(b)** (applied to $q = k$) yields $c_{u,v}(k) = k - 1 \in \{k-1, k\}$. Thus, Corollary 7.124 is proven in Case 1.

Let us now consider Case 2. In this case, we don't have $k \in \{u, u+1, \ldots, v\}$. In other words, we have $k \notin \{u, u+1, \ldots, v\}$. Combining $k \in [n]$ with $k \notin \{u, u+1, \ldots, v\}$, we obtain $k \in [n] \setminus \{u, u+1, \ldots, v\}$. Hence, Lemma 7.123 **(a)** (applied to $q = k$) yields $c_{u,v}(k) = k \in \{k-1, k\}$. Thus, Corollary 7.124 is proven in Case 2.

We have now proven Corollary 7.124 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Corollary 7.124 always holds. $\square$

Next comes another simple property of the Lehmer code:

Lemma 7.125. Let $n \in \mathbb{N}$. Let $u \in [n]$. Let $\sigma \in S_n$. Assume that

$$(\sigma(q) = q \text{ for each } q \in \{1, 2, \ldots, u-1\}). \tag{881}$$

Then, $u + \ell_u(\sigma) = \sigma(u)$.

Proof of Lemma 7.125. We have

$$\sigma\left(\{1,2,\ldots,u-1\}\right) = \left\{ \underbrace{\sigma\left(q\right)}_{\substack{=q \\ \text{(by (881))}}} \mid q \in \{1,2,\ldots,u-1\} \right\}$$
$$= \{q \mid q \in \{1,2,\ldots,u-1\}\}$$
$$= \{1,2,\ldots,u-1\}. \qquad (882)$$

Next, we claim that $\sigma(u) \geq u$.

[*Proof:* Assume the contrary. Thus, $\sigma(u) < u$. Hence, $\sigma(u) \in \{1,2,\ldots,u-1\}$ (because $\sigma(u)$ is a positive integer). Thus,

$$\sigma(u) \in \{1,2,\ldots,u-1\} = \sigma\left(\{1,2,\ldots,u-1\}\right) \qquad \text{(by (882))}.$$

In other words, there exists some $q \in \{1,2,\ldots,u-1\}$ such that $\sigma(u) = \sigma(q)$. Consider this q. But $\sigma \in S_n$; thus, σ is a permutation of $\{1,2,\ldots,n\}$ (since S_n is the set of all permutations of $\{1,2,\ldots,n\}$). Hence, σ is a bijective map, thus an injective map. Thus, from $\sigma(u) = \sigma(q)$, we obtain $u = q \in \{1,2,\ldots,u-1\}$, whence $u \leq u-1 < u$. This is absurd. Hence, we have obtained a contradiction. Thus, our assumption was false. Thus, $\sigma(u) \geq u$ is proven.]

If a and b are any two nonnegative integers satisfying $a \geq b$, then

$$\{1,2,\ldots,a\} \setminus \{1,2,\ldots,b\} = \{b+1,b+2,\ldots,a\}$$

and thus

$$|\{1,2,\ldots,a\} \setminus \{1,2,\ldots,b\}| = |\{b+1,b+2,\ldots,a\}|$$
$$= a-b \qquad (\text{since } a \geq b). \qquad (883)$$

But $\sigma(u) \geq 1$ (since $\sigma(u) \in [n]$) and $u \geq 1$ (since $u \in [n]$). Hence, $\sigma(u)-1$ and $u-1$ are nonnegative integers. These integers satisfy $\sigma(u)-1 \geq u-1$ (since $\sigma(u) \geq u$).

Hence, (883) (applied to $a = \sigma(u)-1$ and $b = u-1$) yields

$$|\{1,2,\ldots,\sigma(u)-1\} \setminus \{1,2,\ldots,u-1\}| = (\sigma(u)-1)-(u-1) = \sigma(u)-u.$$

But Lemma 5.48 **(b)** (applied to $i = u$) yields

$$\begin{aligned}
\ell_u(\sigma) &= \left| \underbrace{[\sigma(u)-1]}_{\substack{=\{1,2,\ldots,\sigma(u)-1\} \\ \text{(by the definition of } [\sigma(u)-1])}} \setminus \sigma \left(\underbrace{[u-1]}_{\substack{=\{1,2,\ldots,u-1\} \\ \text{(by the definition of } [u-1])}} \right) \right| \\
&= \left| \{1,2,\ldots,\sigma(u)-1\} \setminus \underbrace{\sigma(\{1,2,\ldots,u-1\})}_{\substack{=\{1,2,\ldots,u-1\} \\ \text{(by (882))}}} \right| \\
&= |\{1,2,\ldots,\sigma(u)-1\} \setminus \{1,2,\ldots,u-1\}| \\
&= \sigma(u) - u.
\end{aligned}$$

In other words, $u + \ell_u(\sigma) = \sigma(u)$. This proves Lemma 7.125. □

Next, we state a lemma that will be crucial for our solution:

Lemma 7.126. Let $n \in \mathbb{N}$. Let u and v be two elements of $[n]$ such that $u \leq v$. Let $\sigma \in S_n$ be such that $\sigma(u) = v$. Let $\tau \in S_n$ be such that $\tau = (c_{u,v})^{-1} \circ \sigma$. Then:

(a) We have $\tau(u) = u$.

(b) We have $\ell_i(\tau) = \ell_i(\sigma)$ for each $i \in \{u+1, u+2, \ldots, n\}$.

(c) If we have

$$(\sigma(q) = q \text{ for each } q \in \{1,2,\ldots,u-1\}), \tag{884}$$

then we have

$$(\tau(q) = q \text{ for each } q \in \{1,2,\ldots,u\}). \tag{885}$$

Proof of Lemma 7.126. We have

$$c_{u,v} \circ \underbrace{\tau}_{=(c_{u,v})^{-1} \circ \sigma} = \underbrace{c_{u,v} \circ (c_{u,v})^{-1}}_{=\text{id}} \circ \sigma = \sigma. \tag{886}$$

Also, Lemma 7.123 **(c)** yields $c_{u,v}(u) = v$. Hence, $(c_{u,v})^{-1}(v) = u$. Now,

$$\underbrace{\tau}_{=(c_{u,v})^{-1} \circ \sigma}(u) = \left((c_{u,v})^{-1} \circ \sigma\right)(u) = (c_{u,v})^{-1} \left(\underbrace{\sigma(u)}_{=v} \right) = (c_{u,v})^{-1}(v) = u.$$

This proves Lemma 7.126 **(a)**.

Recall that S_n is the set of all permutations of $\{1,2,\ldots,n\}$. In other words, S_n is the set of all permutations of $[n]$ (since $[n] = \{1,2,\ldots,n\}$). We have $\sigma \in S_n$. In other words, σ is a permutation of $[n]$ (since S_n is the set of all permutations of $[n]$). In other words, σ is a bijective map $[n] \to [n]$. The same argument (applied to τ instead of σ) shows that τ is a bijective map $[n] \to [n]$. The maps σ and τ are bijective, and thus injective.

(b) First, let us prove the following auxiliary claim:

> *Claim 1:* Let $k \in [n]$ be such that $k \neq u$. Then, $\tau(k) - 1 \leq \sigma(k) \leq \tau(k)$.

[*Proof of Claim 1:* If we had $\tau(k) = \tau(u)$, then we would have $k = u$ (since τ is injective), which would contradict $k \neq u$. Thus, we cannot have $\tau(k) = \tau(u)$. Hence, we have $\tau(k) \neq \tau(u) = u$ (by Lemma 7.126 **(a)**).

Hence, Corollary 7.124 (applied to $\tau(k)$ instead of k) yields $c_{u,v}(\tau(k)) \in \{\tau(k) - 1, \tau(k)\}$. Now,

$$\underbrace{\sigma}_{\substack{=c_{u,v}\circ\tau \\ \text{(by (886))}}}(k) = (c_{u,v} \circ \tau)(k) = c_{u,v}(\tau(k)) \in \{\tau(k) - 1, \tau(k)\}$$
$$= \{\tau(k) - 1, \tau(k), \ldots, \tau(k)\}.$$

Hence, $\tau(k) - 1 \leq \sigma(k) \leq \tau(k)$. This proves Claim 1.]

Now, let $i \in \{u+1, u+2, \ldots, n\}$. We shall now show the following claim:

> *Claim 2:* Let $j \in \{i+1, i+2, \ldots, n\}$. Then, we have the following logical equivalence:
> $$(\tau(i) > \tau(j)) \iff (\sigma(i) > \sigma(j)).$$

[*Proof of Claim 2:* We have $j \in \{i+1, i+2, \ldots, n\}$, thus $j \geq i+1 > i$, and therefore $j \neq i$.

We have $i \in \{u+1, u+2, \ldots, n\}$, thus $i \geq u+1 > u$, and thus $i \neq u$. Now, $j > i > u$, thus $j \neq u$.

Also, $i \in \{u+1, u+2, \ldots, n\} \subseteq [n]$ and $j \in \{i+1, i+2, \ldots, n\} \subseteq [n]$ (since $i \in [n]$).

Claim 1 (applied to $k = i$) yields $\tau(i) - 1 \leq \sigma(i) \leq \tau(i)$ (since $i \in [n]$ and $i \neq u$).

Claim 1 (applied to $k = j$) yields $\tau(j) - 1 \leq \sigma(j) \leq \tau(j)$ (since $j \in [n]$ and $j \neq u$).

The map τ is injective. Thus, if we had $\tau(j) = \tau(i)$, then we would have $j = i$, which would contradict $j \neq i$. Hence, we cannot have $\tau(j) = \tau(i)$. Thus, we have $\tau(j) \neq \tau(i)$. In other words, $\tau(i) \neq \tau(j)$. The same argument (but applied to σ instead of τ) yields $\sigma(i) \neq \sigma(j)$.

Now, we have the logical implication

$$(\tau(i) > \tau(j)) \implies (\sigma(i) > \sigma(j)). \tag{887}$$

[*Proof of (887):* Assume that $\tau(i) > \tau(j)$ holds. We must prove that $\sigma(i) > \sigma(j)$ holds.

We have $\tau(i) > \tau(j)$, thus $\tau(i) \geq \tau(j) + 1$ (since $\tau(i)$ and $\tau(j)$ are integers). Hence, $\tau(i) - 1 \geq \tau(j)$.

We have $\tau(i)-1 \leq \sigma(i)$, thus $\sigma(i) \geq \tau(i)-1 \geq \tau(j) \geq \sigma(j)$ (since $\sigma(j) \leq \tau(j)$). Combining this with $\sigma(i) \neq \sigma(j)$, we obtain $\sigma(i) > \sigma(j)$.

Now, forget our assumption that $\tau(i) > \tau(j)$. We thus have proven that if $\tau(i) > \tau(j)$ holds, then $\sigma(i) > \sigma(j)$ holds. Thus, the implication (887) is proven.]

We also have the logical implication

$$(\sigma(i) > \sigma(j)) \implies (\tau(i) > \tau(j)). \tag{888}$$

[*Proof of (888):* Assume that $\sigma(i) > \sigma(j)$ holds. We must prove that $\tau(i) > \tau(j)$ holds.

We have $\sigma(i) > \sigma(j)$, thus $\sigma(i) \geq \sigma(j)+1$ (since $\sigma(i)$ and $\sigma(j)$ are integers). Hence, $\sigma(i)-1 \geq \sigma(j)$.

We have $\tau(j)-1 \leq \sigma(j)$, thus $\sigma(j) \geq \tau(j)-1$ and therefore $\sigma(j)+1 \geq \tau(j)$.

But from $\sigma(i) \leq \tau(i)$, we obtain $\tau(i) \geq \sigma(i) \geq \sigma(j)+1 \geq \tau(j)$. Combining this with $\tau(i) \neq \tau(j)$, we obtain $\tau(i) > \tau(j)$.

Now, forget our assumption that $\sigma(i) > \sigma(j)$. We thus have proven that if $\sigma(i) > \sigma(j)$ holds, then $\tau(i) > \tau(j)$ holds. Thus, the implication (888) is proven.]

Combining the two implications (887) and (888), we obtain the logical equivalence $(\tau(i) > \tau(j)) \iff (\sigma(i) > \sigma(j))$. This proves Claim 2.]

Now, recall that $\ell_i(\sigma)$ is the number of all $j \in \{i+1, i+2, \ldots, n\}$ such that $\sigma(i) > \sigma(j)$ (by the definition of $\ell_i(\sigma)$). In other words,

$$\ell_i(\sigma) = |\{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}|. \tag{889}$$

The same argument (applied to τ instead of σ) yields

$$\ell_i(\tau) = |\{j \in \{i+1, i+2, \ldots, n\} \mid \tau(i) > \tau(j)\}|. \tag{890}$$

But

$$\left\{ j \in \{i+1, i+2, \ldots, n\} \mid \underbrace{\tau(i) > \tau(j)}_{\substack{\iff (\sigma(i) > \sigma(j)) \\ \text{(by Claim 2)}}} \right\} = \{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}.$$

Hence, (890) becomes

$$\ell_i(\tau) = \left| \underbrace{\{j \in \{i+1, i+2, \ldots, n\} \mid \tau(i) > \tau(j)\}}_{=\{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}} \right| = |\{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j)\}| = \ell_i(\sigma)$$

(by (889)). This proves Lemma 7.126 **(b)**.

(c) Assume that we have (884). We must prove that we have (885). In other words, we must prove that $\tau(q) = q$ for each $q \in \{1, 2, \ldots, u\}$.

So let $q \in \{1, 2, \ldots, u\}$. We then must prove that $\tau(q) = q$.

We are in one of the following two cases:

Case 1: We have $q = u$.

Case 2: We have $q \neq u$.

Let us first consider Case 1. In this case, we have $q = u$. Thus,

$$\tau\left(\underbrace{q}_{=u}\right) = \tau(u) = u \qquad \text{(by Lemma 7.126 **(a)**)}$$
$$= q.$$

Hence, $\tau(q) = q$ is proven in Case 1.

Let us now consider Case 2. In this case, we have $q \neq u$. Combining $q \in \{1, 2, \ldots, u\}$ with $q \neq u$, we obtain $q \in \{1, 2, \ldots, u\} \setminus \{u\} = \{1, 2, \ldots, u-1\}$. Thus, (884) yields $\sigma(q) = q$. Also, from $q \in \{1, 2, \ldots, u-1\}$, we obtain $q \leq u - 1 < u$.

Also,

$$q \in \{1, 2, \ldots, u\} \subseteq \{1, 2, \ldots, n\} \qquad (\text{since } u \leq n \text{ (because } u \in [n] = \{1, 2, \ldots, n\}))$$
$$= [n].$$

If we had $q \in \{u, u+1, \ldots, v\}$, then we would have $q \geq u$, which would contradict $q < u$. Hence, we cannot have $q \in \{u, u+1, \ldots, v\}$. In other words, we have $q \notin \{u, u+1, \ldots, v\}$.

Combining $q \in [n]$ with $q \notin \{u, u+1, \ldots, v\}$, we obtain $q \in [n] \setminus \{u, u+1, \ldots, v\}$. Hence, Lemma 7.123 **(a)** yields $c_{u,v}(q) = q$.

Now,

$$\underbrace{\tau}_{=(c_{u,v})^{-1} \circ \sigma}(q) = \left((c_{u,v})^{-1} \circ \sigma\right)(q) = (c_{u,v})^{-1}\left(\underbrace{\sigma(q)}_{=q}\right) = (c_{u,v})^{-1}(q) = q$$

(since $c_{u,v}(q) = q$). Hence, $\tau(q) = q$ is proven in Case 2.

We have now proven $\tau(q) = q$ in each of the two Cases 1 and 2. We thus conclude that $\tau(q) = q$ always holds.

Thus, we have shown that we have (885). This proves Lemma 7.126 **(c)**. □

Lemma 7.127. Let $n \in \mathbb{N}$. Define the permutation $w_0 \in S_n$ as in Exercise 5.1 **(c)**. Then,

$$\ell_i(w_0) = n - i \qquad \text{for each } i \in \{1, 2, \ldots, n\}. \tag{891}$$

Proof of Lemma 7.127. Let $i \in \{1, 2, \ldots, n\}$. Thus, $i \geq 1$ and $i \leq n$ and $i \in \{1, 2, \ldots, n\} = [n]$ (since $[n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$)).

We know that $\ell_i(w_0)$ is the number of all $j \in \{i+1, i+2, \ldots, n\}$ such that $w_0(i) > w_0(j)$ (by the definition of $\ell_i(w_0)$). In other words,

$$\ell_i(w_0) = |\{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\}|. \qquad (892)$$

Now, let $k \in \{i+1, i+2, \ldots, n\}$. Thus, $k \geq i+1 > i$. The definition of w_0 yields $w_0(i) = n+1-i$ and $w_0(k) = n+1-\underbrace{k}_{>i} < n+1-i = w_0(i)$, so that $w_0(i) > w_0(k)$. Hence, k is a $j \in \{i+1, i+2, \ldots, n\}$ satisfying $w_0(i) > w_0(j)$. In other words, $k \in \{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\}$.

Now, forget that we fixed k. We thus have proven that $k \in \{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\}$ for each $k \in \{i+1, i+2, \ldots, n\}$. In other words, we have

$$\{i+1, i+2, \ldots, n\} \subseteq \{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\}.$$

Combining this inclusion with the (obvious) inclusion

$$\{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\} \subseteq \{i+1, i+2, \ldots, n\},$$

we obtain

$$\{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\} = \{i+1, i+2, \ldots, n\}.$$

Hence, (892) becomes

$$\ell_i(w_0) = \left| \underbrace{\{j \in \{i+1, i+2, \ldots, n\} \mid w_0(i) > w_0(j)\}}_{=\{i+1, i+2, \ldots, n\}} \right| = |\{i+1, i+2, \ldots, n\}|$$
$$= n - i \qquad (\text{since } i \leq n).$$

This proves Lemma 7.127. ☐

Next, let us state an induction principle that will come useful in our solution of Exercise 5.21 **(c)**:

Theorem 7.128. Let $g \in \mathbb{Z}$ and $h \in \mathbb{Z}$. For each $p \in \{g, g+1, \ldots, h\}$, let $\mathcal{A}(p)$ be a logical statement.

Assume the following:

Assumption 1: If $g \leq h$, then the statement $\mathcal{A}(h)$ holds.

Assumption 2: If $u \in \{g+1, g+2, \ldots, h\}$ is such that $\mathcal{A}(u)$ holds, then $\mathcal{A}(u-1)$ also holds.

Then, $\mathcal{A}(p)$ holds for each $p \in \{g, g+1, \ldots, h\}$.

Proof of Theorem 7.128. Theorem 7.128 is exactly Theorem 2.162, except that some names have been changed:

- The variable n has been renamed as p.
- The variable m has been renamed as u.

Thus, Theorem 7.128 holds (since Theorem 2.162 holds). □

We shall only need the particular case of Theorem 7.128 in which $g = 0$ and $h = n$:

Corollary 7.129. Let $n \in \mathbb{Z}$. For each $p \in \{0, 1, \ldots, n\}$, let $\mathcal{A}(p)$ be a logical statement.

Assume the following:

Assumption 1: If $0 \leq n$, then the statement $\mathcal{A}(n)$ holds.

Assumption 2: If $u \in \{1, 2, \ldots, n\}$ is such that $\mathcal{A}(u)$ holds, then $\mathcal{A}(u-1)$ also holds.

Then, $\mathcal{A}(p)$ holds for each $p \in \{0, 1, \ldots, n\}$.

Proof of Corollary 7.129. Corollary 7.129 is just the particular case of Theorem 7.128 obtained by taking $g = 0$ and $h = n$. □

At this point we have proven enough auxiliary facts to render the solution of Exercise 5.21 fairly anodyne:

Solution to Exercise 5.21. Let us first observe that the definition of $c_{u,v}$ given in Exercise 5.21 is exactly the definition of $c_{u,v}$ given in Definition 7.104. Thus, there is no conflict of notation.

(a) Let $i \in [n]$. Lemma 7.122 **(a)** yields that the permutation $c_{i,i+\ell_i(\sigma)} \in S_n$ is well-defined. Thus, the permutation $a_i \in S_n$ is well-defined (since a_i was defined by $a_i = c_{i,i+\ell_i(\sigma)}$). This solves Exercise 5.21 **(a)**.

(b) Let $i \in [n]$. Set $i' = i + \ell_i(\sigma)$. We must show that $a_i = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$.

The definition of a_i yields $a_i = c_{i,i+\ell_i(\sigma)} = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$ (by Lemma 7.122 **(b)**). This solves Exercise 5.21 **(b)**.

(c) Forget that we fixed σ (but let us leave n fixed).

Let us define some further notations. If $\sigma \in S_n$ and $i \in [n]$ are arbitrary, then we define a permutation $a_i^{(\sigma)} \in S_n$ by

$$a_i^{(\sigma)} = c_{i,i+\ell_i(\sigma)}. \tag{893}$$

(This is well-defined, because Lemma 7.122 **(a)** yields that the permutation $c_{i,i+\ell_i(\sigma)} \in S_n$ is well-defined.) Thus, $a_i^{(\sigma)}$ is precisely what was called a_i in Exercise 5.21, but we have renamed it $a_i^{(\sigma)}$ in order to make its dependence on σ explicit. (This allows us to simultaneously consider $a_i^{(\sigma)}$ and $a_i^{(\tau)}$ for two different permutations σ and τ.)

If $p \in \{0, 1, \ldots, n\}$ and $\sigma \in S_n$, then we say that σ is *p-lazy* if we have

$$(\sigma(q) = q \text{ for each } q \in \{1, 2, \ldots, p\}).$$

Note that

$$\text{every permutation } \sigma \in S_n \text{ is 0-lazy} \tag{894}$$

[425]. On the other hand,

$$\text{if a permutation } \sigma \in S_n \text{ is } n\text{-lazy, then } \sigma = \text{id} \tag{895}$$

[426].

For each $p \in \{0, 1, \ldots, n\}$, let us define a logical statement $\mathcal{A}(p)$ as follows:

Statement $\mathcal{A}(p)$: Every p-lazy permutation $\sigma \in S_n$ satisfies

$$\sigma = a_{p+1}^{(\sigma)} \circ a_{p+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)}.$$

Our goal is to prove that $\mathcal{A}(p)$ holds for each $p \in \{0, 1, \ldots, n\}$. Once this is done, we will immediately conclude that $\mathcal{A}(0)$ holds, which means that Exercise 5.21 **(c)** holds (since the claim of Exercise 5.21 **(c)** is more or less obviously equivalent to $\mathcal{A}(0)$).

In order to prove that $\mathcal{A}(p)$ holds for each $p \in \{0, 1, \ldots, n\}$, we shall use Corollary 7.129. In order to do so, we need to prove that Assumptions 1 and 2 of Corollary 7.129 hold for these statements $\mathcal{A}(0), \mathcal{A}(1), \ldots, \mathcal{A}(n)$. Let us therefore prove that these two assumptions hold:

[*Proof of Assumption 1:* Assume that $0 \leq n$. (This is clearly true anyway, since $n \in \mathbb{N}$.)

Let $\sigma \in S_n$ be an n-lazy permutation. Thus, $\sigma = \text{id}$ (by (895)). Comparing this with

$$a_{n+1}^{(\sigma)} \circ a_{n+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} = (\text{an empty composition of permutations in } S_n) = \text{id},$$

[425] *Proof of (894):* Let $\sigma \in S_n$. We must prove that σ is 0-lazy.

The statement $(\sigma(q) = q$ for each $q \in \{1, 2, \ldots, 0\})$ is vacuously true (since there exists no $q \in \{1, 2, \ldots, 0\}$ (because $\{1, 2, \ldots, 0\} = \varnothing$)), and therefore is true. In other words, σ is 0-lazy (because σ is 0-lazy if and only if we have $(\sigma(q) = q$ for each $q \in \{1, 2, \ldots, 0\})$ (by the definition of "0-lazy")). This proves (894).

[426] *Proof of (895):* Let $\sigma \in S_n$ be a permutation that is n-lazy. We must prove that $\sigma = \text{id}$.

We have $\sigma \in S_n$. In other words, σ is a permutation of $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\}$). In other words, σ is a bijection from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$.

But σ is n-lazy if and only if we have $(\sigma(q) = q$ for each $q \in \{1, 2, \ldots, n\})$ (by the definition of "n-lazy"). Hence, we have $(\sigma(q) = q$ for each $q \in \{1, 2, \ldots, n\})$ (since σ is n-lazy). Thus, for each $q \in \{1, 2, \ldots, n\}$, we have $\sigma(q) = q = \text{id}(q)$. Hence, $\sigma = \text{id}$ (since both σ and id are maps from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$). This proves (895).

we obtain $\sigma = a_{n+1}^{(\sigma)} \circ a_{n+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)}$.

Now, forget that we fixed σ. We thus have shown that every n-lazy permutation $\sigma \in S_n$ satisfies $\sigma = a_{n+1}^{(\sigma)} \circ a_{n+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)}$. But this is precisely the statement $\mathcal{A}(n)$. Hence, we have shown that the statement $\mathcal{A}(n)$ holds. This concludes the proof of Assumption 1.]

[*Proof of Assumption 2:* Let $u \in \{1,2,\ldots,n\}$ be such that $\mathcal{A}(u)$ holds. We must prove that $\mathcal{A}(u-1)$ also holds.

We have $u \in \{1,2,\ldots,n\} = [n]$.

We have assumed that $\mathcal{A}(u)$ holds. In other words,

$$\text{every } u\text{-lazy permutation } \sigma \in S_n \text{ satisfies } \sigma = a_{u+1}^{(\sigma)} \circ a_{u+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} \tag{896}$$

(because this is what the statement $\mathcal{A}(u)$ says).

Now, we will try to prove that $\mathcal{A}(u-1)$ holds.

Let $\sigma \in S_n$ be a $(u-1)$-lazy permutation. Thus, σ is $(u-1)$-lazy. In other words,

$$(\sigma(q) = q \text{ for each } q \in \{1,2,\ldots,u-1\}) \tag{897}$$

[427]. Hence, Lemma 7.125 yields $u + \ell_u(\sigma) = \sigma(u)$.

Define $v \in [n]$ by $v = \sigma(u)$. (This is well-defined, since $\sigma(u) \in [n]$.) Then, $u + \ell_u(\sigma) = \sigma(u) = v$ (since $v = \sigma(u)$).

From $u + \ell_u(\sigma) = v$, we obtain $v = u + \underbrace{\ell_u(\sigma)}_{\geq 0} \geq u$, so that $u \leq v$.

Now, the definition of $a_u^{(\sigma)}$ yields

$$a_u^{(\sigma)} = c_{u,u+\ell_u(\sigma)} = c_{u,v} \qquad (\text{since } u + \ell_u(\sigma) = v). \tag{898}$$

Define a permutation $\tau \in S_n$ by $\tau = (c_{u,v})^{-1} \circ \sigma$. Hence, Lemma 7.126 **(b)** yields that

$$\ell_i(\tau) = \ell_i(\sigma) \qquad \text{for each } i \in \{u+1, u+2, \ldots, n\}. \tag{899}$$

Thus, each $i \in \{u+1, u+2, \ldots, n\}$ satisfies

$$\begin{aligned} a_i^{(\tau)} &= c_{i,i+\ell_i(\tau)} && \left(\text{by the definition of } a_i^{(\tau)}\right) \\ &= c_{i,i+\ell_i(\sigma)} && (\text{since } \ell_i(\tau) = \ell_i(\sigma) \text{ (by (899))}) \\ &= a_i^{(\sigma)} && (\text{by (893)}). \end{aligned}$$

In other words, $\left(a_{u+1}^{(\tau)}, a_{u+2}^{(\tau)}, \ldots, a_n^{(\tau)}\right) = \left(a_{u+1}^{(\sigma)}, a_{u+2}^{(\sigma)}, \ldots, a_n^{(\sigma)}\right)$. Hence,

$$a_{u+1}^{(\tau)} \circ a_{u+2}^{(\tau)} \circ \cdots \circ a_n^{(\tau)} = a_{u+1}^{(\sigma)} \circ a_{u+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)}. \tag{900}$$

[427] because σ is $(u-1)$-lazy if and only if $(\sigma(q) = q$ for each $q \in \{1,2,\ldots,u-1\})$ (by the definition of "$(u-1)$-lazy")

Recall that $(\sigma(q) = q \text{ for each } q \in \{1, 2, \ldots, u-1\})$. Hence, Lemma 7.126 **(c)** yields that $(\tau(q) = q \text{ for each } q \in \{1, 2, \ldots, u\})$. In other words, the permutation τ is u-lazy[428]. Therefore, (896) (applied to τ instead of σ) yields

$$\tau = a_{u+1}^{(\tau)} \circ a_{u+2}^{(\tau)} \circ \cdots \circ a_n^{(\tau)} = a_{u+1}^{(\sigma)} \circ a_{u+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} \tag{901}$$

(by (900)).

Now, $c_{u,v} \circ \underbrace{\tau}_{=(c_{u,v})^{-1} \circ \sigma} = \underbrace{c_{u,v} \circ (c_{u,v})^{-1}}_{=\mathrm{id}} \circ \sigma = \sigma$, so that

$$\begin{aligned} \sigma &= \underbrace{c_{u,v}}_{\substack{=a_u^{(\sigma)} \\ \text{(by (898))}}} \circ \underbrace{\tau}_{\substack{=a_{u+1}^{(\sigma)} \circ a_{u+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} \\ \text{(by (901))}}} = a_u^{(\sigma)} \circ \left(a_{u+1}^{(\sigma)} \circ a_{u+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} \right) \\ &= a_u^{(\sigma)} \circ a_{u+1}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} = a_{(u-1)+1}^{(\sigma)} \circ a_{(u-1)+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} \end{aligned}$$

(since $u = (u-1) + 1$ and $u + 1 = (u-1) + 2$).

Now, forget that we fixed σ. We thus have proven that

$$\text{every } (u-1)\text{-lazy permutation } \sigma \in S_n \text{ satisfies } \sigma = a_{(u-1)+1}^{(\sigma)} \circ a_{(u-1)+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)}.$$

But this is exactly the statement $\mathcal{A}(u-1)$. Thus, we have proven that the statement $\mathcal{A}(u-1)$ holds. This proves Assumption 2.]

We have now verified that both Assumptions 1 and 2 of Corollary 7.129 hold. Hence, Corollary 7.129 shows that $\mathcal{A}(p)$ holds for each $p \in \{0, 1, \ldots, n\}$. Applying this to $p = 0$, we conclude that $\mathcal{A}(0)$ holds (since $0 \in \{0, 1, \ldots, n\}$). In other words,

$$\text{every } 0\text{-lazy permutation } \sigma \in S_n \text{ satisfies } \sigma = a_{0+1}^{(\sigma)} \circ a_{0+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} \tag{902}$$

(because this is what the statement $\mathcal{A}(0)$ says).

Now, let $\sigma \in S_n$. Let us use the notation a_i (for $i \in [n]$) as defined in Exercise 5.21. The permutation $\sigma \in S_n$ is 0-lazy (by (894)). Thus, (902) yields

$$\sigma = a_{0+1}^{(\sigma)} \circ a_{0+2}^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} = a_1^{(\sigma)} \circ a_2^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)}.$$

But each $i \in [n]$ satisfies

$$\begin{aligned} a_i^{(\sigma)} &= c_{i, i+\ell_i(\sigma)} && \text{(by (893))} \\ &= a_i && \left(\text{since } a_i = c_{i, i+\ell_i(\sigma)} \text{ (by the definition of } a_i) \right). \end{aligned}$$

In other words, each $i \in \{1, 2, \ldots, n\}$ satisfies $a_i^{(\sigma)} = a_i$ (since $[n] = \{1, 2, \ldots, n\}$). In other words, $\left(a_1^{(\sigma)}, a_2^{(\sigma)}, \ldots, a_n^{(\sigma)} \right) = (a_1, a_2, \ldots, a_n)$. Hence, $a_1^{(\sigma)} \circ a_2^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} =$

[428]because τ is u-lazy if and only if $(\tau(q) = q \text{ for each } q \in \{1, 2, \ldots, u\})$ (by the definition of "u-lazy")

$a_1 \circ a_2 \circ \cdots \circ a_n$. Thus, $\sigma = a_1^{(\sigma)} \circ a_2^{(\sigma)} \circ \cdots \circ a_n^{(\sigma)} = a_1 \circ a_2 \circ \cdots \circ a_n$. This solves Exercise 5.21 **(c)**.

(d) *Second solution to Exercise 5.2 **(e)**:* In the following, a *simple transposition* shall mean a permutation of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$). Thus, s_k is a simple transposition for each $k \in \{1, 2, \ldots, n-1\}$.

For each $i \in [n]$, we define a permutation $a_i \in S_n$ as in Exercise 5.21. Then, we claim that each $i \in \{1, 2, \ldots, n\}$ satisfies

$$a_i = (\text{a composition of } \ell_i(\sigma) \text{ simple transpositions}) \qquad (903)$$

(this equality is supposed to mean that a_i is a composition of i simple transpositions; it does not mean that **every** composition of i simple transpositions is a_i).

[*Proof of (903):* Let $i \in \{1, 2, \ldots, n\}$. Thus, $i \in \{1, 2, \ldots, n\} = [n]$. Set $i' = i + \ell_i(\sigma)$. Then, Exercise 5.21 **(b)** yields $a_i = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$.

We have $i' = i + \underbrace{\ell_i(\sigma)}_{\geq 0} \geq i$. Hence, the permutation $s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i$ is a composition of $i' - i$ simple transpositions (since $s_{i'-1}, s_{i'-2}, \ldots, s_i$ are simple transpositions[429]). In view of $s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i = a_i$ and $\underbrace{i'}_{=i+\ell_i(\sigma)} - i = i + \ell_i(\sigma) - i = \ell_i(\sigma)$, this rewrites as follows: The permutation a_i is a composition of $\ell_i(\sigma)$ simple transpositions. In other words, $a_i = (\text{a composition of } \ell_i(\sigma) \text{ simple transpositions})$. This proves (903).]

Proposition 5.46 yields $\ell(\sigma) = \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma)$, thus $\ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma) = \ell(\sigma)$.

Now, Exercise 5.21 **(c)** yields

$$\begin{aligned}
\sigma &= a_1 \circ a_2 \circ \cdots \circ a_n \\
&= (\text{a composition of } \ell_1(\sigma) \text{ simple transpositions}) \\
&\qquad \circ (\text{a composition of } \ell_2(\sigma) \text{ simple transpositions}) \\
&\qquad \circ \cdots \\
&\qquad \circ (\text{a composition of } \ell_n(\sigma) \text{ simple transpositions}) \\
&\qquad \left(\begin{array}{c} \text{since } a_i = (\text{a composition of } \ell_i(\sigma) \text{ simple transpositions}) \\ \text{for each } i \in \{1, 2, \ldots, n\} \text{ (by (903))} \end{array}\right) \\
&= \left(\text{a composition of } \underbrace{\ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma)}_{=\ell(\sigma)} \text{ simple transpositions}\right) \\
&= (\text{a composition of } \ell(\sigma) \text{ simple transpositions}).
\end{aligned}$$

In other words, σ is a composition of $\ell(\sigma)$ simple transpositions. In other words, σ is a composition of $\ell(\sigma)$ permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$)

[429] since s_k is a simple transposition for each $k \in \{1, 2, \ldots, n-1\}$

(because the simple transpositions are precisely the permutations of the form s_k (with $k \in \{1, 2, \ldots, n-1\}$)). Thus, Exercise 5.2 **(e)** is solved again.

(e) *Second solution to Exercise 5.1* ***(c)****:* Define $\sigma \in S_n$ by $\sigma = w_0$. For each $i \in [n]$, we define a permutation $a_i \in S_n$ as in Exercise 5.21. Then,

$$a_i = s_{n-1} \circ s_{n-2} \circ \cdots \circ s_i \qquad \text{for each } i \in \{1, 2, \ldots, n\}. \tag{904}$$

[*Proof of (904):* Let $i \in \{1, 2, \ldots, n\}$. Thus, $i \in \{1, 2, \ldots, n\} = [n]$.

From $\sigma = w_0$, we obtain $\ell_i(\sigma) = \ell_i(w_0) = n - i$ (by (891)). In other words, $i + \ell_i(\sigma) = n$.

Set $i' = i + \ell_i(\sigma)$. Then, Exercise 5.21 **(b)** yields

$$a_i = s_{i'-1} \circ s_{i'-2} \circ \cdots \circ s_i = s_{n-1} \circ s_{n-2} \circ \cdots \circ s_i$$

(since $i' = i + \ell_i(\sigma) = n$). This proves (904).]

Now, from $\sigma = w_0$, we get

$$\begin{aligned} w_0 = \sigma &= a_1 \circ a_2 \circ \cdots \circ a_n \qquad \text{(by Exercise 5.21 } \textbf{(c)}\text{)} \\ &= (s_{n-1} \circ s_{n-2} \circ \cdots \circ s_1) \circ (s_{n-1} \circ s_{n-2} \circ \cdots \circ s_2) \circ \cdots \circ (s_{n-1} \circ s_{n-2} \circ \cdots \circ s_n) \\ &\qquad \left(\begin{array}{c} \text{since } a_i = s_{n-1} \circ s_{n-2} \circ \cdots \circ s_i \text{ for each } i \in \{1, 2, \ldots, n\} \\ \text{(by (904))} \end{array} \right). \end{aligned}$$

[430] This is an explicit way to write w_0 as a composition of several permutations of the form s_i (with $i \in \{1, 2, \ldots, n-1\}$). Thus, Exercise 5.1 **(c)** is solved again.

[*Remark:* The way of writing w_0 that we have just obtained is **not** the exact same way that we found in our first solution to Exercise 5.1 **(c)** above. Indeed, the latter way is

$$w_0 = (s_0 \circ s_{-1} \circ \cdots \circ s_1) \circ (s_1 \circ s_0 \circ \cdots \circ s_1) \circ \cdots \circ (s_{n-1} \circ s_{n-2} \circ \cdots \circ s_1).$$

It is, however, possible to derive one of these two representations of w_0 from the other.] □

7.61. Solution to Exercise 5.22

In this section, we shall use the notations introduced in Definition 5.54 and in Definition 5.55.

We begin with the (extremely simple) proof of Proposition 5.56:

Proof of Proposition 5.56. We just need to prove that $\begin{cases} \sigma(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \in X$ for each $x \in X$. But this is true, because:

[430] Note that the factor $(s_{n-1} \circ s_{n-2} \circ \cdots \circ s_n)$ is an empty composition of permutations in S_n, and thus equals id. We can omit this factor (when $n > 0$).

- If $x \in Y$, then $\begin{cases} \sigma(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} = \sigma(x) \in Y \subseteq X$.
- Otherwise, $\begin{cases} \sigma(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} = x \in X$.

Thus, Proposition 5.56 is proven. □

Proposition 5.56 justifies Definition 5.55.

Proof of Proposition 5.58. **(a)** Let $\alpha : Y \to Y$ and $\beta : Y \to Y$ be two maps. Let $x \in X$. We shall prove that

$$(\alpha \circ \beta)^{(Y \to X)}(x) = \left(\alpha^{(Y \to X)} \circ \beta^{(Y \to X)}\right)(x). \tag{905}$$

[*Proof of (905):* We are in one of the following two cases:

Case 1: We have $x \in Y$.

Case 2: We don't have $x \in Y$.

Let us first consider Case 1. In this case, we have $x \in Y$. The definition of $\beta^{(Y \to X)}$ yields

$$\beta^{(Y \to X)}(x) = \begin{cases} \beta(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} = \beta(x) \qquad (\text{since } x \in Y),$$

so that $\beta^{(Y \to X)}(x) = \beta\left(\underbrace{x}_{\in Y}\right) \in \beta(Y) \subseteq Y$ (since β is a map from Y to Y).

Now,

$$\begin{aligned}
\left(\alpha^{(Y \to X)} \circ \beta^{(Y \to X)}\right)(x) &= \alpha^{(Y \to X)}\left(\underbrace{\beta^{(Y \to X)}(x)}_{=\beta(x)}\right) = \alpha^{(Y \to X)}(\beta(x)) \\
&= \begin{cases} \alpha(\beta(x)), & \text{if } \beta(x) \in Y; \\ \beta(x), & \text{if } \beta(x) \notin Y \end{cases} \qquad \left(\text{by the definition of } \alpha^{(Y \to X)}\right) \\
&= \alpha(\beta(x)) \qquad (\text{since } \beta(x) \in Y) \\
&= (\alpha \circ \beta)(x).
\end{aligned}$$

Compared with

$$\begin{aligned}
(\alpha \circ \beta)^{(Y \to X)}(x) &= \begin{cases} (\alpha \circ \beta)(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \left(\text{by the definition of } (\alpha \circ \beta)^{(Y \to X)}\right) \\
&= (\alpha \circ \beta)(x) \qquad (\text{since } x \in Y),
\end{aligned}$$

this yields $(\alpha \circ \beta)^{(Y \to X)}(x) = \left(\alpha^{(Y \to X)} \circ \beta^{(Y \to X)}\right)(x)$. We thus have proven (905) in Case 1.

Let us now consider Case 2. In this case, we don't have $x \in Y$. In other words, we have $x \notin Y$. The definition of $\beta^{(Y\to X)}$ yields

$$\beta^{(Y\to X)}(x) = \begin{cases} \beta(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} = x \qquad (\text{since } x \notin Y).$$

Now,

$$\begin{aligned} \left(\alpha^{(Y\to X)} \circ \beta^{(Y\to X)}\right)(x) &= \alpha^{(Y\to X)}\left(\underbrace{\beta^{(Y\to X)}(x)}_{=x}\right) = \alpha^{(Y\to X)}(x) \\ &= \begin{cases} \alpha(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \left(\text{by the definition of } \alpha^{(Y\to X)}\right) \\ &= x \qquad (\text{since } x \notin Y). \end{aligned}$$

Compared with

$$\begin{aligned} (\alpha \circ \beta)^{(Y\to X)}(x) &= \begin{cases} (\alpha \circ \beta)(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \left(\text{by the definition of } (\alpha \circ \beta)^{(Y\to X)}\right) \\ &= x \qquad (\text{since } x \notin Y), \end{aligned}$$

this yields $(\alpha \circ \beta)^{(Y\to X)}(x) = \left(\alpha^{(Y\to X)} \circ \beta^{(Y\to X)}\right)(x)$. We thus have proven (905) in Case 2.

Now, we have proven the equality (905) in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, this yields that (905) always holds.]

Now, forget that we fixed x. We thus have proven that $(\alpha \circ \beta)^{(Y\to X)}(x) = \left(\alpha^{(Y\to X)} \circ \beta^{(Y\to X)}\right)(x)$ for each $x \in X$. In other words, $(\alpha \circ \beta)^{(Y\to X)} = \alpha^{(Y\to X)} \circ \beta^{(Y\to X)}$. This proves Proposition 5.58 **(a)**.

(b) Each $x \in X$ satisfies

$$\begin{aligned} (\mathrm{id}_Y)^{(Y\to X)}(x) &= \begin{cases} \mathrm{id}_Y(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \left(\text{by the definition of } (\mathrm{id}_Y)^{(Y\to X)}\right) \\ &= \begin{cases} x, & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad (\text{since } \mathrm{id}_Y(x) = x \text{ whenever } x \in Y) \\ &= x = \mathrm{id}_X(x). \end{aligned}$$

In other words, $(\mathrm{id}_Y)^{(Y\to X)} = \mathrm{id}_X$. This proves Proposition 5.58 **(b)**.

(c) Let $\sigma \in S_Y$. Thus, σ is a permutation of Y (since S_Y is the set of all permutations of Y). In other words, σ is a bijective map from Y to Y. Hence, the map $\sigma^{(Y\to X)} : X \to X$ is well-defined. Also, the map σ is bijective, thus invertible.

Hence, its inverse σ^{-1} is also a well-defined map from Y to Y. Therefore, the map $\left(\sigma^{-1}\right)^{(Y\to X)} : X \to X$ is well-defined.

Proposition 5.58 **(a)** (applied to $\alpha = \sigma$ and $\beta = \sigma^{-1}$) yields

$$\left(\sigma \circ \sigma^{-1}\right)^{(Y\to X)} = \sigma^{(Y\to X)} \circ \left(\sigma^{-1}\right)^{(Y\to X)}.$$

Hence,

$$\sigma^{(Y\to X)} \circ \left(\sigma^{-1}\right)^{(Y\to X)} = \Bigg(\underbrace{\sigma \circ \sigma^{-1}}_{=\mathrm{id}_Y}\Bigg)^{(Y\to X)} = \left(\mathrm{id}_Y\right)^{(Y\to X)} = \mathrm{id}_X \qquad (906)$$

(by Proposition 5.58 **(b)**). Also, Proposition 5.58 **(a)** (applied to $\alpha = \sigma^{-1}$ and $\beta = \sigma$) yields

$$\left(\sigma^{-1} \circ \sigma\right)^{(Y\to X)} = \left(\sigma^{-1}\right)^{(Y\to X)} \circ \sigma^{(Y\to X)}.$$

Hence,

$$\left(\sigma^{-1}\right)^{(Y\to X)} \circ \sigma^{(Y\to X)} = \Bigg(\underbrace{\sigma^{-1} \circ \sigma}_{=\mathrm{id}_Y}\Bigg)^{(Y\to X)} = \left(\mathrm{id}_Y\right)^{(Y\to X)} = \mathrm{id}_X \qquad (907)$$

(by Proposition 5.58 **(b)**).

Combining (906) with (907), we conclude that the maps $\sigma^{(Y\to X)}$ and $\left(\sigma^{-1}\right)^{(Y\to X)}$ are mutually inverse. Thus, the map $\sigma^{(Y\to X)}$ is invertible, and hence bijective. Hence, $\sigma^{(Y\to X)}$ is a bijective map from X to X. In other words, $\sigma^{(Y\to X)}$ is a permutation of X. In other words, $\sigma^{(Y\to X)} \in S_X$ (since S_X is the set of all permutations of X). Furthermore, it satisfies $\left(\sigma^{-1}\right)^{(Y\to X)} = \left(\sigma^{(Y\to X)}\right)^{-1}$ (since the maps $\sigma^{(Y\to X)}$ and $\left(\sigma^{-1}\right)^{(Y\to X)}$ are mutually inverse). Thus, Proposition 5.58 **(c)** is proven.

(d) First, we claim that

$$\left\{\delta^{(Y\to X)} \mid \delta \in S_Y\right\} \subseteq \left\{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\right\}. \qquad (908)$$

[*Proof of (908):* Let $\alpha \in \left\{\delta^{(Y\to X)} \mid \delta \in S_Y\right\}$. Thus, $\alpha = \delta^{(Y\to X)}$ for some map $\delta \in S_Y$. Consider this δ. Proposition 5.58 **(c)** (applied to $\sigma = \delta$) yields that $\delta^{(Y\to X)} \in S_X$ and $\left(\delta^{-1}\right)^{(Y\to X)} = \left(\delta^{(Y\to X)}\right)^{-1}$.

Now, for every $z \in X \setminus Y$, we have

$$\underbrace{\alpha}_{=\delta^{(Y\to X)}}(z) = \delta^{(Y\to X)}(z) = \begin{cases} \delta(z), & \text{if } z \in Y; \\ z, & \text{if } z \notin Y \end{cases} \qquad \left(\text{by the definition of } \delta^{(Y\to X)}\right)$$
$$= z \qquad (\text{since } z \notin Y \text{ (because } z \in X \setminus Y)).$$

Hence, we have shown that $\alpha(z) = z$ for every $z \in X \setminus Y$. Thus, α is a $\tau \in S_X$ that satisfies $\tau(z) = z$ for every $z \in X \setminus Y$ (since $\alpha = \delta^{(Y \to X)} \in S_X$). In other words,

$$\alpha \in \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}.$$

Since we have proven this for **each** $\alpha \in \left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\}$, we thus conclude that $\left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\} \subseteq \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}$. This proves (908).]

Next, we claim that

$$\{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\} \subseteq \left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\}. \qquad (909)$$

[*Proof of (909):* Let $\eta \in \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}$. Thus, η is an element of S_X and satisfies

$$\eta(z) = z \text{ for every } z \in X \setminus Y. \qquad (910)$$

We have $\eta \in S_X$. In other words, η is a permutation of X. In other words, η is a bijective map from X to X. Hence, the map η is bijective, and thus both injective and surjective.

Let $y \in Y$. Assume (for the sake of contradiction) that $\eta(y) \notin Y$. Hence, $\eta(y) \in X \setminus Y$. Thus, (910) (applied to $z = \eta(y)$) yields $\eta(\eta(y)) = \eta(y)$. Since η is injective, this yields $\eta(y) = y$, hence $y = \eta(y) \in X \setminus Y$, so that $y \notin Y$, which contradicts $y \in Y$. Hence, our assumption (that $\eta(y) \notin Y$) was wrong. Hence, we have $\eta(y) \in Y$.

Now, forget that we fixed y. We thus have shown that every $y \in Y$ satisfies $\eta(y) \in Y$. Hence, we can define a map $\varepsilon : Y \to Y$ by

$$(\varepsilon(y) = \eta(y) \qquad \text{for each } y \in Y). \qquad (911)$$

Consider this ε.

This map ε is injective[431] and surjective[432]. Hence, this map ε is bijective. Thus, ε is a bijection from Y to Y. In other words, ε is a permutation of the set Y. In other words, $\varepsilon \in S_Y$.

[431] *Proof.* Let $a \in Y$ and $b \in Y$ be such that $\varepsilon(a) = \varepsilon(b)$. By the definition of ε, we have $\varepsilon(a) = \eta(a)$ and $\varepsilon(b) = \eta(b)$. Now, $\eta(a) = \varepsilon(a) = \varepsilon(b) = \eta(b)$. Hence, $a = b$ (since the map η is injective).

Now, forget that we fixed a and b. We thus have shown that every $a \in Y$ and $b \in Y$ satisfying $\varepsilon(a) = \varepsilon(b)$ must satisfy $a = b$. In other words, the map ε is injective.

[432] *Proof.* Let $w \in Y$. Then, $w \in Y \subseteq X = \eta(X)$ (since η is surjective). In other words, there exists some $z \in X$ such that $w = \eta(z)$. Consider this z. If we had $z \in X \setminus Y$, then we would have

$$\eta(z) = z \qquad (\text{by (910)}) \\ \notin Y \qquad (\text{since } z \in X \setminus Y),$$

which would contradict $\eta(z) = w \in Y$. Thus, we cannot have $z \in X \setminus Y$. Hence, we must have $z \notin X \setminus Y$. Combining $z \in X$ with $z \notin X \setminus Y$, we obtain $z \in X \setminus (X \setminus Y) \subseteq Y$. Hence, $\varepsilon(z)$ is

Now, let $x \in X$. We are going to prove that $\varepsilon^{(Y \to X)}(x) = \eta(x)$.

Let us rewrite the value $\eta(x)$ depending on whether x belongs to Y or not:

- If $x \in Y$, then $\eta(x) = \varepsilon(x)$ (because the definition of ε yields $\varepsilon(x) = \eta(x)$).
- If $x \notin Y$, then $\eta(x) = x$ (because $x \notin Y$ entails $x \in X \setminus Y$, and therefore (910) (applied to $z = x$) yields $\eta(x) = x$).

Combining these two observations, we obtain

$$\eta(x) = \begin{cases} \varepsilon(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases}.$$

But the definition of $\varepsilon^{(Y \to X)}$ yields

$$\varepsilon^{(Y \to X)}(x) = \begin{cases} \varepsilon(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases}.$$

Comparing the preceding two equations, we obtain $\varepsilon^{(Y \to X)}(x) = \eta(x)$.

Now, forget that we fixed x. We now have proven that $\varepsilon^{(Y \to X)}(x) = \eta(x)$ for every $x \in X$. In other words, $\varepsilon^{(Y \to X)} = \eta$. Hence, $\eta = \varepsilon^{(Y \to X)} \in \left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\}$ (since $\varepsilon \in S_Y$).

Now, forget that we fixed η. We thus have shown that every $\eta \in \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}$ satisfies $\eta \in \left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\}$. In other words, we have

$$\{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\} \subseteq \left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\}.$$

This proves (909).]

Combining the two inclusions (908) and (909), we obtain

$$\left\{\delta^{(Y \to X)} \mid \delta \in S_Y\right\} = \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}.$$

This proves Proposition 5.58 **(d)**.

(e) Define a subset G of S_X by

$$G = \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\}. \qquad (912)$$

---

well-defined. The definition of ε yields $\varepsilon(z) = \eta(z) = w$. Hence, $w = \varepsilon\left(\underbrace{z}_{\in Y}\right) \in \varepsilon(Y)$.

Now, forget that we fixed w. We thus have shown that $w \in \varepsilon(Y)$ for each $w \in Y$. In other words, $Y \subseteq \varepsilon(Y)$. In other words, ε is surjective.

Proposition 5.58 **(d)** yields

$$\left\{\delta^{(Y\to X)} \mid \delta\in S_Y\right\} = \{\tau\in S_X \mid \tau(z)=z \text{ for every } z\in X\setminus Y\}$$
$$= G \qquad (\text{by (912)}). \tag{913}$$

Thus, $\left\{\delta^{(Y\to X)} \mid \delta\in S_Y\right\} = G\subseteq G$. In other words, we have $\delta^{(Y\to X)}\in G$ for each $\delta\in S_Y$. Hence, we can define a map $R:S_Y\to G$ by

$$\left(R(\delta)=\delta^{(Y\to X)} \qquad \text{for each } \delta\in S_Y\right).$$

Consider this map R.

The map R is injective[433] and surjective[434]. Hence, the map R is bijective.

Now, the map R is a map from S_Y to G such that $\left(R(\delta)=\delta^{(Y\to X)} \text{ for each } \delta\in S_Y\right)$. In view of (912), this rewrites as follows: The map R is a map from S_Y to $\{\tau\in S_X \mid \tau(z)=z \text{ for every } z\in X\setminus Y\}$ such that $\left(R(\delta)=\delta^{(Y\to X)} \text{ for each } \delta\in S_Y\right)$. Hence, R is the map

$$S_Y\to\{\tau\in S_X \mid \tau(z)=z \text{ for every } z\in X\setminus Y\},$$
$$\delta\mapsto\delta^{(Y\to X)}.$$

[433] *Proof.* Let $\delta\in S_Y$ and $\varepsilon\in S_Y$ be such that $R(\delta)=R(\varepsilon)$. We shall prove that $\delta=\varepsilon$.

Let $y\in Y$. Thus, $y\in Y\subseteq X$. The definition of R yields $R(\delta)=\delta^{(Y\to X)}$ and $R(\varepsilon)=\varepsilon^{(Y\to X)}$. Hence, $\delta^{(Y\to X)}=R(\delta)=R(\varepsilon)=\varepsilon^{(Y\to X)}$.

But $y\in X$. Thus, the definition of $\delta^{(Y\to X)}$ yields

$$\delta^{(Y\to X)}(y)=\begin{cases}\delta(y), & \text{if } y\in Y;\\ y, & \text{if } y\notin Y\end{cases} = \delta(y) \qquad (\text{since } y\in Y).$$

The same argument (applied to ε instead of δ) yields $\varepsilon^{(Y\to X)}(y)=\varepsilon(y)$. Now, from $\delta^{(Y\to X)}=\varepsilon^{(Y\to X)}$, we obtain $\delta^{(Y\to X)}(y)=\varepsilon^{(Y\to X)}(y)=\varepsilon(y)$. Comparing this with $\delta^{(Y\to X)}(y)=\delta(y)$, we obtain $\delta(y)=\varepsilon(y)$.

Now, forget that we fixed y. We thus have shown that $\delta(y)=\varepsilon(y)$ for each $y\in Y$. In other words, $\delta=\varepsilon$.

Now, forget that we fixed δ and ε. We thus have proven that if $\delta\in S_Y$ and $\varepsilon\in S_Y$ are such that $R(\delta)=R(\varepsilon)$, then $\delta=\varepsilon$. In other words, the map R is injective.

[434] *Proof.* Let $\eta\in G$. Then,

$$\eta\in G=\left\{\delta^{(Y\to X)} \mid \delta\in S_Y\right\} \qquad (\text{by (912)}).$$

In other words, there exists some $\delta\in S_Y$ such that $\eta=\delta^{(Y\to X)}$. Consider this δ.

Now, the definition of $R(\delta)$ yields $R(\delta)=\delta^{(Y\to X)}$. Comparing this with $\eta=\delta^{(Y\to X)}$, we obtain $\eta=R\left(\underbrace{\delta}_{\in S_Y}\right)\in R(S_Y)$.

Now, forget that we fixed η. We thus have shown that $\eta\in R(S_Y)$ for each $\eta\in G$. In other words, $G\subseteq R(S_Y)$. In other words, the map R is surjective.

Thus, the map

$$S_Y \to \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus Y\},$$
$$\delta \mapsto \delta^{(Y\to X)}$$

is well-defined and bijective (since R is bijective). This proves Proposition 5.58 **(e)**. □

Proof of Proposition 5.59. Let $x \in X$. We are going to show that $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) = \sigma^{(Z\to X)}(x)$.

The definition of $\sigma^{(Z\to X)}$ yields

$$\sigma^{(Z\to X)}(x) = \begin{cases} \sigma(x), & \text{if } x \in Z; \\ x, & \text{if } x \notin Z \end{cases}. \tag{914}$$

Now, we distinguish between two cases:

Case 1: We have $x \in Y$.

Case 2: We don't have $x \in Y$.

Let us first consider Case 1. In this case, we have $x \in Y$. The definition of $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}$ yields

$$\begin{aligned}\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) &= \begin{cases} \sigma^{(Z\to Y)}(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} = \sigma^{(Z\to Y)}(x) \qquad (\text{since } x \in Y) \\ &= \begin{cases} \sigma(x), & \text{if } x \in Z; \\ x, & \text{if } x \notin Z \end{cases} \qquad \left(\text{by the definition of } \sigma^{(Z\to Y)}\right) \\ &= \sigma^{(Z\to X)}(x) \qquad (\text{by (914)}).\end{aligned}$$

We thus have shown that $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) = \sigma^{(Z\to X)}(x)$ in Case 1.

Let us now consider Case 2. In this case, we don't have $x \in Y$. In other words, we have $x \notin Y$. Thus, $x \notin Z$ (since otherwise, we would have $x \in Z \subseteq Y$, contradicting $x \notin Y$). Thus, (914) becomes

$$\sigma^{(Z\to X)}(x) = \begin{cases} \sigma(x), & \text{if } x \in Z; \\ x, & \text{if } x \notin Z \end{cases} = x \qquad (\text{since } x \notin Z). \tag{915}$$

Now, the definition of $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}$ yields

$$\begin{aligned}\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) &= \begin{cases} \sigma^{(Z\to Y)}(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} = x \qquad (\text{since } x \notin Y) \\ &= \sigma^{(Z\to X)}(x) \qquad (\text{by (915)}).\end{aligned}$$

We thus have shown that $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) = \sigma^{(Z\to X)}(x)$ in Case 2.

We now have proven that $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) = \sigma^{(Z\to X)}(x)$ holds in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, this yields that $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) = \sigma^{(Z\to X)}(x)$ always holds.

Now, forget that we fixed x. We thus have shown that $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)}(x) = \sigma^{(Z\to X)}(x)$ for every $x \in X$. In other words, $\left(\sigma^{(Z\to Y)}\right)^{(Y\to X)} = \sigma^{(Z\to X)}$. This proves Proposition 5.59. □

Proof of Proposition 5.60. We have $X \setminus (X \setminus Y) = Y$ (since $Y \subseteq X$).

Define a map $\eta : X \to X$ by $\eta = \alpha^{(Y\to X)}$.

Define a map $\zeta : X \to X$ by $\zeta = \beta^{(X\setminus Y\to X)}$.

Let now $x \in X$. We are going to prove that $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$.

We distinguish between two cases:

Case 1: We have $x \in Y$.

Case 2: We don't have $x \in Y$.

Let us first consider Case 1. In this case, we have $x \in Y$. Thus, $x \in Y = X \setminus (X \setminus Y)$. Hence, $x \in X$ and $x \notin X \setminus Y$. Now, recall that $\zeta = \beta^{(X\setminus Y\to X)}$. Hence,

$$\begin{aligned} \zeta(x) = \beta^{(X\setminus Y\to X)}(x) &= \begin{cases} \beta(x), & \text{if } x \in X \setminus Y; \\ x, & \text{if } x \notin X \setminus Y \end{cases} \qquad \left(\text{by the definition of } \beta^{(X\setminus Y\to X)}\right) \\ &= x \qquad (\text{since } x \notin X \setminus Y). \end{aligned}$$

Also, $\eta = \alpha^{(Y\to X)}$, so that

$$\begin{aligned} \eta(x) = \alpha^{(Y\to X)}(x) &= \begin{cases} \alpha(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \left(\text{by the definition of } \alpha^{(Y\to X)}\right) \\ &= \alpha\left(\underbrace{x}_{\in Y}\right) \qquad (\text{since } x \in Y) \\ &\in \alpha(Y) \subseteq Y \qquad (\text{since } \alpha \text{ is a map } Y \to Y) \\ &= X \setminus (X \setminus Y), \end{aligned}$$

and thus $\eta(x) \notin X \setminus Y$. Now,

$$(\eta \circ \zeta)(x) = \eta\left(\underbrace{\zeta(x)}_{=x}\right) = \eta(x).$$

Compared with

$$\begin{aligned}(\zeta \circ \eta)(x) &= \underbrace{\zeta}_{=\beta^{(X\setminus Y\to X)}}(\eta(x)) = \beta^{(X\setminus Y\to X)}(\eta(x)) \\ &= \begin{cases} \beta(\eta(x)), & \text{if } \eta(x) \in X \setminus Y; \\ \eta(x), & \text{if } \eta(x) \notin X \setminus Y \end{cases} \qquad \left(\text{by the definition of } \beta^{(X\setminus Y\to X)}\right) \\ &= \eta(x) \qquad (\text{since } \eta(x) \notin X \setminus Y),\end{aligned}$$

this yields $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$. Thus, $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$ is proven in Case 1.

Now, let us consider Case 2. In this case, we don't have $x \in Y$. Thus, $x \notin Y$. Combined with $x \in X$, this yields $x \in X \setminus Y$. Now, recall that $\zeta = \beta^{(X\setminus Y\to X)}$, so that

$$\begin{aligned}\zeta(x) &= \beta^{(X\setminus Y\to X)}(x) \\ &= \begin{cases} \beta(x), & \text{if } x \in X \setminus Y; \\ x, & \text{if } x \notin X \setminus Y \end{cases} \qquad \left(\text{by the definition of } \beta^{(X\setminus Y\to X)}\right) \\ &= \beta\left(\underbrace{x}_{\in X\setminus Y}\right) \qquad (\text{since } x \in X \setminus Y) \\ &\in \beta(X \setminus Y) \subseteq X \setminus Y \qquad (\text{since } \beta \text{ is a map } X \setminus Y \to X \setminus Y),\end{aligned}$$

and thus $\zeta(x) \notin Y$. Also, $\eta = \alpha^{(Y\to X)}$, so that

$$\begin{aligned}\eta(x) &= \alpha^{(Y\to X)}(x) = \begin{cases} \alpha(x), & \text{if } x \in Y; \\ x, & \text{if } x \notin Y \end{cases} \qquad \left(\text{by the definition of } \alpha^{(Y\to X)}\right) \\ &= x \qquad (\text{since } x \notin Y).\end{aligned}$$

Now,

$$\begin{aligned}(\eta \circ \zeta)(x) &= \eta(\zeta(x)) = \alpha^{(Y\to X)}(\zeta(x)) \qquad \left(\text{since } \eta = \alpha^{(Y\to X)}\right) \\ &= \begin{cases} \alpha(\zeta(x)), & \text{if } \zeta(x) \in Y; \\ \zeta(x), & \text{if } \zeta(x) \notin Y \end{cases} \qquad \left(\text{by the definition of } \alpha^{(Y\to X)}\right) \\ &= \zeta(x) \qquad (\text{since } \zeta(x) \notin Y).\end{aligned}$$

Compared with

$$(\zeta \circ \eta)(x) = \zeta\left(\underbrace{\eta(x)}_{=x}\right) = \zeta(x),$$

this yields $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$. Thus, $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$ is proven in Case 2.

We now have proven the equality $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$ in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, this yields that the equality $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$ always holds.

Now, forget that we fixed x. We thus have proven that $(\eta \circ \zeta)(x) = (\zeta \circ \eta)(x)$ for every $x \in X$. In other words, $\eta \circ \zeta = \zeta \circ \eta$. In view of $\eta = \alpha^{(Y \to X)}$ and $\zeta = \beta^{(X \setminus Y \to X)}$, this rewrites as

$$\alpha^{(Y \to X)} \circ \beta^{(X \setminus Y \to X)} = \beta^{(X \setminus Y \to X)} \circ \alpha^{(Y \to X)}.$$

This proves Proposition 5.60. □

Proof of Proposition 5.62. We will prove Proposition 5.62 by strong induction over $|X|$:

Induction step: Let $N \in \mathbb{N}$. Assume that Proposition 5.62 holds whenever $|X| < N$. We need to prove that Proposition 5.62 holds whenever $|X| = N$.

We have assumed that Proposition 5.62 holds whenever $|X| < N$. In other words, the following claim holds:

> *Claim 1:* Let X be a finite set satisfying $|X| < N$. Let $x \in X$. Let $\pi \in S_X$. Then, there exists a $\sigma \in \left\{ \delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}} \right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$.

We need to prove that Proposition 5.62 holds whenever $|X| = N$. In other words, we need to prove the following claim:

> *Claim 2:* Let X be a finite set satisfying $|X| = N$. Let $x \in X$. Let $\pi \in S_X$. Then, there exists a $\sigma \in \left\{ \delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}} \right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$.

Before we start proving Claim 2, let us first establish an auxiliary claim, which follows easily from Claim 1:

> *Claim 3:* Let X be a finite set satisfying $|X| < N$. Let $\pi \in S_X$. Then, there exists a $\sigma \in S_X$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$.

[*Proof of Claim 3:* If $X = \varnothing$, then Claim 3 holds for simple reasons[435]. Hence, for the rest of this proof of Claim 3, we can WLOG assume that we don't have $X = \varnothing$. Assume this.

There exists some $x \in X$ (since we don't have $X = \varnothing$). Consider this x.

[435] *Proof.* Assume that $X = \varnothing$. We must prove that Claim 3 holds.

We have $X = \varnothing$. Thus, no $x \in X$ exists. Hence, $\left(\pi \circ \pi \circ \pi^{-1}\right)(x) = \pi^{-1}(x)$ holds for all $x \in X$ (indeed, this is vacuously true, since no $x \in X$ exists). In other words, we have $\pi \circ \pi \circ \pi^{-1} = \pi^{-1}$. Thus, there exists a $\sigma \in S_X$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$ (namely, $\sigma = \pi$). Hence, Claim 3 holds.

We thus have proven that if $X = \varnothing$, then Claim 3 holds. Qed.

Claim 1 shows that there exists a $\sigma \in \left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$. Consider this σ, and denote it by α. Thus, α is an element of $\left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$ and satisfies $\alpha \circ \pi \circ \alpha^{-1} = \pi^{-1}$.

We know that α is an element of $\left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$. In other words, there exists some $\delta \in S_{X\setminus\{x\}}$ such that $\alpha = \delta^{(X\setminus\{x\}\to X)}$. Consider this δ.

Now, Proposition 5.58 **(c)** (applied to $Y = X \setminus \{x\}$ and $\sigma = \delta$) yields that δ satisfies $\delta^{(X\setminus\{x\}\to X)} \in S_X$ and $\left(\delta^{-1}\right)^{(X\setminus\{x\}\to X)} = \left(\delta^{(X\setminus\{x\}\to X)}\right)^{-1}$. Now, $\alpha = \delta^{(X\setminus\{x\}\to X)} \in S_X$ and $\alpha \circ \pi \circ \alpha^{-1} = \pi^{-1}$. Hence, there exists a $\sigma \in S_X$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$ (namely, $\sigma = \alpha$). This proves Claim 3.]

For the sake of convenience, we shall also derive a simple consequence of Proposition 5.58:

Claim 4: Let X be a set. Let $x \in X$. Let $\gamma \in S_{X\setminus\{x\}}$ and $\varepsilon \in S_{X\setminus\{x\}}$. Then:

(a) We have $\gamma^{(X\setminus\{x\}\to X)} \in S_X$.

(b) We have $(\gamma \circ \varepsilon)^{(X\setminus\{x\}\to X)} = \gamma^{(X\setminus\{x\}\to X)} \circ \varepsilon^{(X\setminus\{x\}\to X)}$.

(c) We have

$$\left(\gamma \circ \varepsilon \circ \gamma^{-1}\right)^{(X\setminus\{x\}\to X)} = \gamma^{(X\setminus\{x\}\to X)} \circ \varepsilon^{(X\setminus\{x\}\to X)} \circ \left(\gamma^{(X\setminus\{x\}\to X)}\right)^{-1}.$$

[*Proof of Claim 4:* Clearly, γ and ε are elements of $S_{X\setminus\{x\}}$, thus permutations of the set $X \setminus \{x\}$ (since $S_{X\setminus\{x\}}$ is the set of all permutations of the set $X \setminus \{x\}$). In other words, γ and ε are bijective maps from $X \setminus \{x\}$ to $X \setminus \{x\}$. Thus, their inverses γ^{-1} and ε^{-1} are maps from $X \setminus \{x\}$ to $X \setminus \{x\}$ as well.

Proposition 5.58 **(c)** (applied to $Y = X \setminus \{x\}$ and $\sigma = \gamma$) shows that γ satisfies $\gamma^{(X\setminus\{x\}\to X)} \in S_X$ and $\left(\gamma^{-1}\right)^{(X\setminus\{x\}\to X)} = \left(\gamma^{(X\setminus\{x\}\to X)}\right)^{-1}$. This proves Claim 4 **(a)**. Also, Proposition 5.58 **(a)** (applied to $X \setminus \{x\}$, γ and ε instead of Y, α and β) yields $(\gamma \circ \varepsilon)^{(X\setminus\{x\}\to X)} = \gamma^{(X\setminus\{x\}\to X)} \circ \varepsilon^{(X\setminus\{x\}\to X)}$. This proves Claim 4 **(b)**. Now, Proposition 5.58 **(a)** (applied to $X \setminus \{x\}$, $\gamma \circ \varepsilon$ and γ^{-1} instead of Y, α and β) yields

$$\begin{aligned}\left(\gamma \circ \varepsilon \circ \gamma^{-1}\right)^{(X\setminus\{x\}\to X)} &= \underbrace{(\gamma \circ \varepsilon)^{(X\setminus\{x\}\to X)}}_{=\gamma^{(X\setminus\{x\}\to X)} \circ \varepsilon^{(X\setminus\{x\}\to X)}} \circ \underbrace{\left(\gamma^{-1}\right)^{(X\setminus\{x\}\to X)}}_{=\left(\gamma^{(X\setminus\{x\}\to X)}\right)^{-1}} \\ &= \gamma^{(X\setminus\{x\}\to X)} \circ \varepsilon^{(X\setminus\{x\}\to X)} \circ \left(\gamma^{(X\setminus\{x\}\to X)}\right)^{-1}.\end{aligned}$$

This proves Claim 4 **(c)**.]

We are now ready to prove Claim 2:

[*Proof of Claim 2:* Since $x \in X$, we have $|X \setminus \{x\}| = |X| - 1 < |X| = N$.

Applying Proposition 5.58 **(d)** to $Y = X \setminus \{x\}$, we obtain

$$\left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$$
$$= \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus (X \setminus \{x\})\}. \tag{916}$$

We know that $\pi \in S_X$. In other words, π is a permutation of X (since S_X is the set of all permutations of X). In other words, π is a bijective map $X \to X$.

We are in one of the following two cases:

Case 1: We have $\pi(x) = x$.

Case 2: We don't have $\pi(x) = x$.

Let us first consider Case 1. In this case, we have $\pi(x) = x$. Thus, $\pi(z) = z$ for every $z \in X \setminus (X \setminus \{x\})$ [436]. Hence, π is a $\tau \in S_X$ satisfying $(\tau(z) = z \text{ for every } z \in X \setminus (X \setminus \{x\}))$. In other words, $\pi \in \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus (X \setminus \{x\})\}$. Hence,

$$\begin{aligned}\pi &\in \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus (X \setminus \{x\})\} \\ &= \left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\} \qquad \text{(by (916))} \\ &= \left\{\left(\pi'\right)^{(X\setminus\{x\}\to X)} \mid \pi' \in S_{X\setminus\{x\}}\right\}\end{aligned}$$

(here, we have renamed the index δ as π'). In other words, there exists a $\pi' \in S_{X\setminus\{x\}}$ such that $\pi = (\pi')^{(X\setminus\{x\}\to X)}$. Consider this π'.

Proposition 5.58 **(c)** (applied to $Y = X \setminus \{x\}$ and $\sigma = \pi'$) shows that π' satisfies $(\pi')^{(X\setminus\{x\}\to X)} \in S_X$ and

$$\left(\left(\pi'\right)^{-1}\right)^{(X\setminus\{x\}\to X)} = \left(\underbrace{\left(\pi'\right)^{(X\setminus\{x\}\to X)}}_{=\pi}\right)^{-1} = \pi^{-1}.$$

Now, Claim 3 (applied to $X \setminus \{x\}$ and π' instead of X and π) yields that there exists a $\sigma \in S_{X\setminus\{x\}}$ such that $\sigma \circ \pi' \circ \sigma^{-1} = (\pi')^{-1}$. Denote this σ by σ'. Thus, $\sigma' \in S_{X\setminus\{x\}}$ and $\sigma' \circ \pi' \circ (\sigma')^{-1} = (\pi')^{-1}$.

Define the map $\sigma'' : X \to X$ by

$$\sigma'' = \left(\sigma'\right)^{(X\setminus\{x\}\to X)}.$$

Thus, $\sigma'' = \delta^{(X\setminus\{x\}\to X)}$ for some $\delta \in S_{X\setminus\{x\}}$ (namely, for $\delta = \sigma'$). In other words,

$$\sigma'' \in \left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}. \tag{917}$$

[436] *Proof.* Let $z \in X \setminus (X \setminus \{x\})$. Then, $z \in X \setminus (X \setminus \{x\}) \subseteq \{x\}$, so that $z = x$. Hence, $\pi(z) = \pi(x) = x = z$, qed.

Now, Claim 4 **(c)** (applied to $\gamma = \sigma'$ and $\varepsilon = \pi'$) yields

$$\left(\sigma' \circ \pi' \circ (\sigma')^{-1}\right)^{(X\setminus\{x\}\to X)}$$
$$= \left(\underbrace{(\sigma')^{(X\setminus\{x\}\to X)}}_{=\sigma''}\right) \circ \left(\underbrace{(\pi')^{(X\setminus\{x\}\to X)}}_{=\pi}\right) \circ \left(\underbrace{(\sigma')^{(X\setminus\{x\}\to X)}}_{=\sigma''}\right)^{-1}$$
$$= \sigma'' \circ \pi \circ (\sigma'')^{-1}.$$

Compared with

$$\left(\underbrace{\sigma' \circ \pi' \circ (\sigma')^{-1}}_{=(\pi')^{-1}}\right)^{(X\setminus\{x\}\to X)} = \left((\pi')^{-1}\right)^{(X\setminus\{x\}\to X)} = \pi^{-1},$$

this yields $\sigma'' \circ \pi \circ (\sigma'')^{-1} = \pi^{-1}$. So we have shown that σ'' is an element of $\left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$ (by (917)) and satisfies $\sigma'' \circ \pi \circ (\sigma'')^{-1} = \pi^{-1}$. Hence, there exists a $\sigma \in \left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$ (namely, $\sigma = \sigma''$). Thus, Claim 2 holds in Case 1.

Now, let us consider Case 2. In this case, we don't have $\pi(x) = x$. Hence, we have $\pi(x) \neq x$.

Let $y = \pi(x)$. Combining $y = \pi(x) \in X$ with $y = \pi(x) \neq x$, we obtain $y \in X \setminus \{x\}$. Hence, x and y are two distinct elements of X. Thus, the transposition $t_{x,y}$ of X is well-defined (according to Definition 5.36). Consider this transposition $t_{x,y}$.

Lemma 7.93 **(a)** (applied to $i = x$ and $j = y$) yields $t_{x,y}(x) = y$. Lemma 7.93 **(b)** (applied to $i = x$ and $j = y$) yields $t_{x,y}(y) = x$. Lemma 7.93 **(d)** (applied to $i = x$ and $j = y$) yields $t_{x,y} \circ t_{x,y} = \mathrm{id}$. Thus, $(t_{x,y})^{-1} = t_{x,y}$. Finally, Lemma 7.93 **(c)** (applied to $i = x$ and $j = y$) shows that we have

$$t_{x,y}(k) = k \text{ for each } k \in X \setminus \{x, y\}. \tag{918}$$

Thus,

$$t_{x,y} \in \left\{\alpha^{(\{x,y\}\to X)} \mid \alpha \in S_{\{x,y\}}\right\} \tag{919}$$

[437]. In other words, there exists some $\alpha \in S_{\{x,y\}}$ such that

$$t_{x,y} = \alpha^{(\{x,y\}\to X)}. \tag{920}$$

[437] *Proof of (919):* We know that $t_{x,y}$ is a permutation of X. Hence, $t_{x,y}$ belongs to S_X (since S_X is the set of all permutations of X) and satisfies $t_{x,y}(z) = z$ for every $z \in X \setminus \{x, y\}$ (by (918), applied to $k = z$). In other words, $t_{x,y}$ is a $\tau \in S_X$ satisfying ($\tau(z) = z$ for every $z \in X \setminus \{x, y\}$). In other words, $t_{x,y} \in \{\tau \in S_X \mid \tau(z) = z \text{ for every } z \in X \setminus \{x, y\}\}$. But Proposition 5.58 **(d)** (applied to

Consider this α. (Of course, we can easily tell what this α is: It is the permutation of $\{x,y\}$ that swaps x with y. But we don't need to know this.)

Both $t_{x,y}$ and π are permutations of X. Thus, their composition $t_{x,y}\circ\pi$ is a permutation of X as well. In other words, $t_{x,y}\circ\pi\in S_X$.

Set $\widetilde{\varepsilon}=t_{x,y}\circ\pi$. Then, $\widetilde{\varepsilon}=t_{x,y}\circ\pi\in S_X$. Also,

$$\underbrace{\widetilde{\varepsilon}}_{=t_{x,y}\circ\pi}(x)=\left(t_{x,y}\circ\pi\right)(x)=t_{x,y}\left(\underbrace{\pi(x)}_{=y}\right)=t_{x,y}(y)=x.$$

Thus, $\widetilde{\varepsilon}(z)=z$ for every $z\in X\setminus(X\setminus\{x\})$ [438]. Hence, $\widetilde{\varepsilon}$ is a $\tau\in S_X$ satisfying $(\tau(z)=z$ for every $z\in X\setminus(X\setminus\{x\}))$. Thus,

$$\begin{aligned}\widetilde{\varepsilon}&\in\{\tau\in S_X\ \mid\ \tau(z)=z\text{ for every }z\in X\setminus(X\setminus\{x\})\}\\&=\left\{\delta^{(X\setminus\{x\}\to X)}\ \mid\ \delta\in S_{X\setminus\{x\}}\right\}\qquad\text{(by (916))}\\&=\left\{\varepsilon^{(X\setminus\{x\}\to X)}\ \mid\ \varepsilon\in S_{X\setminus\{x\}}\right\}\end{aligned}$$

(here, we have renamed the index δ as ε). In other words, there exists some $\varepsilon\in S_{X\setminus\{x\}}$ such that

$$\widetilde{\varepsilon}=\varepsilon^{(X\setminus\{x\}\to X)}.\tag{921}$$

Consider this ε.

Proposition 5.58 **(c)** (applied to $Y=X\setminus\{x\}$ and $\sigma=\varepsilon$) shows that ε satisfies $\varepsilon^{(X\setminus\{x\}\to X)}\in S_X$ and

$$\left(\varepsilon^{-1}\right)^{(X\setminus\{x\}\to X)}=\left(\underbrace{\varepsilon^{(X\setminus\{x\}\to X)}}_{\substack{=\widetilde{\varepsilon}\\\text{(by (921))}}}\right)^{-1}=\widetilde{\varepsilon}^{-1}.\tag{922}$$

Now, Claim 1 (applied to $X\setminus\{x\}$, ε and y instead of X, π and x) yields that there exists a $\sigma\in\left\{\delta^{((X\setminus\{x\})\setminus\{y\}\to X\setminus\{x\})}\ \mid\ \delta\in S_{(X\setminus\{x\})\setminus\{y\}}\right\}$ such that $\sigma\circ\varepsilon\circ\sigma^{-1}=\varepsilon^{-1}$.

$\{x,y\}$ instead of Y) yields

$$\left\{\delta^{(\{x,y\}\to X)}\ \mid\ \delta\in S_{\{x,y\}}\right\}=\{\tau\in S_X\ \mid\ \tau(z)=z\text{ for every }z\in X\setminus\{x,y\}\}.$$

Now,

$$\begin{aligned}t_{x,y}&\in\{\tau\in S_X\ \mid\ \tau(z)=z\text{ for every }z\in X\setminus\{x,y\}\}\\&=\left\{\delta^{(\{x,y\}\to X)}\ \mid\ \delta\in S_{\{x,y\}}\right\}=\left\{\alpha^{(\{x,y\}\to X)}\ \mid\ \alpha\in S_{\{x,y\}}\right\}\end{aligned}$$

(here, we have renamed the index δ as α). This proves (919).

[438] *Proof.* Let $z\in X\setminus(X\setminus\{x\})$. Then, $z\in X\setminus(X\setminus\{x\})\subseteq\{x\}$, so that $z=x$. Hence, $\widetilde{\varepsilon}(z)=\widetilde{\varepsilon}(x)=x=z$, qed.

Denote this σ by γ. Then,

$$\gamma \in \left\{ \delta^{((X\setminus\{x\})\setminus\{y\} \to X\setminus\{x\})} \mid \delta \in S_{(X\setminus\{x\})\setminus\{y\}} \right\} \text{ and } \gamma \circ \varepsilon \circ \gamma^{-1} = \varepsilon^{-1}.$$

We have

$$\begin{aligned} \gamma &\in \left\{ \delta^{((X\setminus\{x\})\setminus\{y\} \to X\setminus\{x\})} \mid \delta \in S_{(X\setminus\{x\})\setminus\{y\}} \right\} \\ &= \left\{ \delta^{(X\setminus\{x,y\} \to X\setminus\{x\})} \mid \delta \in S_{X\setminus\{x,y\}} \right\} \qquad (\text{since } (X\setminus\{x\})\setminus\{y\} = X\setminus\{x,y\}) \\ &= \left\{ \beta^{(X\setminus\{x,y\} \to X\setminus\{x\})} \mid \beta \in S_{X\setminus\{x,y\}} \right\} \end{aligned}$$

(here, we have renamed the index δ as β). In other words, there exists some $\beta \in S_{X\setminus\{x,y\}}$ such that $\gamma = \beta^{(X\setminus\{x,y\} \to X\setminus\{x\})}$. Consider this β.

Clearly, $X \setminus \{x,y\}$ is a subset of $X \setminus \{x\}$ (since $X \setminus \{x,y\} = (X \setminus \{x\}) \setminus \{y\}$). Hence, from Proposition 5.58 **(c)** (applied to $X \setminus \{x\}$, $X \setminus \{x,y\}$ and β instead of X, Y and σ), we obtain that β satisfies $\beta^{(X\setminus\{x,y\} \to X\setminus\{x\})} \in S_{X\setminus\{x\}}$ and $\left(\beta^{-1}\right)^{(X\setminus\{x,y\} \to X\setminus\{x\})} = \left(\beta^{(X\setminus\{x,y\} \to X\setminus\{x\})}\right)^{-1}$. Thus,

$$\gamma = \beta^{(X\setminus\{x,y\} \to X\setminus\{x\})} \in S_{X\setminus\{x\}}.$$

Recall that $\gamma \in S_{X\setminus\{x\}}$. Thus, Proposition 5.58 **(c)** (applied to $Y = X \setminus \{x\}$ and $\sigma = \gamma$) yields that γ satisfies $\gamma^{(X\setminus\{x\} \to X)} \in S_X$ and $\left(\gamma^{-1}\right)^{(X\setminus\{x\} \to X)} = \left(\gamma^{(X\setminus\{x\} \to X)}\right)^{-1}$.

Now, set

$$\widetilde{\gamma} = \gamma^{(X\setminus\{x\} \to X)}. \tag{923}$$

Thus,

$$\begin{aligned} \widetilde{\gamma} &= \gamma^{(X\setminus\{x\} \to X)} = \left(\beta^{(X\setminus\{x,y\} \to X\setminus\{x\})}\right)^{(X\setminus\{x\} \to X)} \qquad \left(\text{since } \gamma = \beta^{(X\setminus\{x,y\} \to X\setminus\{x\})}\right) \\ &= \beta^{(X\setminus\{x,y\} \to X)} \end{aligned} \tag{924}$$

(by Proposition 5.59, applied to $Y = X \setminus \{x\}$ and $Z = X \setminus \{x,y\}$ and $\sigma = \beta$).

Now, Proposition 5.60 (applied to $Y = \{x,y\}$) yields

$$\alpha^{(\{x,y\} \to X)} \circ \beta^{(X\setminus\{x,y\} \to X)} = \beta^{(X\setminus\{x,y\} \to X)} \circ \alpha^{(\{x,y\} \to X)}.$$

In view of (920) and (924), this rewrites as

$$t_{x,y} \circ \widetilde{\gamma} = \widetilde{\gamma} \circ t_{x,y}. \tag{925}$$

Both γ and ε are elements of $S_{X\setminus\{x\}}$. In other words, both γ and ε are permutations of $X \setminus \{x\}$. Hence, $\gamma \circ \varepsilon$ is a permutation of $X \setminus \{x\}$ as well. In other words, $\gamma \circ \varepsilon \in S_{X\setminus\{x\}}$.

Moreover, Claim 4 **(b)** yields

$$(\gamma \circ \varepsilon)^{(X\setminus\{x\}\to X)} = \underbrace{\gamma^{(X\setminus\{x\}\to X)}}_{\substack{=\widetilde{\gamma} \\ \text{(by (923))}}} \circ \underbrace{\varepsilon^{(X\setminus\{x\}\to X)}}_{\substack{=\widetilde{\varepsilon} \\ \text{(by (921))}}} = \widetilde{\gamma} \circ \widetilde{\varepsilon},$$

so that $\widetilde{\gamma} \circ \widetilde{\varepsilon} = (\gamma \circ \varepsilon)^{(X\setminus\{x\}\to X)}$. Hence, $\widetilde{\gamma} \circ \widetilde{\varepsilon} = \delta^{(X\setminus\{x\}\to X)}$ for some $\delta \in S_{X\setminus\{x\}}$ (namely, $\delta = \gamma \circ \varepsilon$), because $\gamma \circ \varepsilon \in S_{X\setminus\{x\}}$. In other words,

$$\widetilde{\gamma} \circ \widetilde{\varepsilon} \in \left\{ \delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}} \right\}. \tag{926}$$

On the other hand, recall that $\gamma \circ \varepsilon \circ \gamma^{-1} = \varepsilon^{-1}$. Thus,

$$\left(\gamma \circ \varepsilon \circ \gamma^{-1}\right)^{(X\setminus\{x\}\to X)} = \left(\varepsilon^{-1}\right)^{(X\setminus\{x\}\to X)} = \widetilde{\varepsilon}^{-1} \qquad \text{(by (922))}.$$

Compared with

$$\begin{aligned} &\left(\gamma \circ \varepsilon \circ \gamma^{-1}\right)^{(X\setminus\{x\}\to X)} \\ &= \left(\underbrace{\gamma^{(X\setminus\{x\}\to X)}}_{\substack{=\widetilde{\gamma} \\ \text{(by (923))}}}\right) \circ \left(\underbrace{\varepsilon^{(X\setminus\{x\}\to X)}}_{\substack{=\widetilde{\varepsilon} \\ \text{(by (921))}}}\right) \circ \left(\underbrace{\gamma^{(X\setminus\{x\}\to X)}}_{\substack{=\widetilde{\gamma} \\ \text{(by (923))}}}\right)^{-1} \qquad \text{(by Claim 4 } \textbf{(c)}\text{)} \\ &= \widetilde{\gamma} \circ \widetilde{\varepsilon} \circ \widetilde{\gamma}^{-1}, \end{aligned}$$

this yields

$$\widetilde{\varepsilon}^{-1} = \widetilde{\gamma} \circ \widetilde{\varepsilon} \circ \widetilde{\gamma}^{-1}. \tag{927}$$

But

$$\begin{aligned} &\left(\underbrace{t_{x,y} \circ \widetilde{\gamma}}_{\substack{=\widetilde{\gamma} \circ t_{x,y} \\ \text{(by (925))}}}\right)^{-1} \circ \left(\widetilde{\gamma} \circ \underbrace{\widetilde{\varepsilon}}_{=t_{x,y} \circ \pi}\right) \\ &= \left(\widetilde{\gamma} \circ t_{x,y}\right)^{-1} \circ \left(\widetilde{\gamma} \circ t_{x,y} \circ \pi\right) = \underbrace{\left(\widetilde{\gamma} \circ t_{x,y}\right)^{-1} \circ \left(\widetilde{\gamma} \circ t_{x,y}\right)}_{=\mathrm{id}} \circ \pi = \pi. \end{aligned}$$

Thus, $\pi = \left(t_{x,y} \circ \widetilde{\gamma}\right)^{-1} \circ (\widetilde{\gamma} \circ \widetilde{\varepsilon})$, so that

$$\begin{aligned}
&(\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ \underbrace{\pi}_{=\left(t_{x,y} \circ \widetilde{\gamma}\right)^{-1} \circ (\widetilde{\gamma} \circ \widetilde{\varepsilon})} \circ (\widetilde{\gamma} \circ \widetilde{\varepsilon})^{-1} \\
&= (\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ \left(t_{x,y} \circ \widetilde{\gamma}\right)^{-1} \circ \underbrace{(\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ (\widetilde{\gamma} \circ \widetilde{\varepsilon})^{-1}}_{=\mathrm{id}} = (\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ \underbrace{\left(t_{x,y} \circ \widetilde{\gamma}\right)^{-1}}_{=\widetilde{\gamma}^{-1} \circ \left(t_{x,y}\right)^{-1}} \\
&= \underbrace{(\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ \widetilde{\gamma}^{-1}}_{\substack{=\widetilde{\gamma} \circ \widetilde{\varepsilon} \circ \widetilde{\gamma}^{-1} = \widetilde{\varepsilon}^{-1} \\ \text{(by (927))}}} \circ \underbrace{\left(t_{x,y}\right)^{-1}}_{=t_{x,y}} = \widetilde{\varepsilon}^{-1} \circ t_{x,y}. \qquad (928)
\end{aligned}$$

Since $\widetilde{\varepsilon} = t_{x,y} \circ \pi$, we have $\widetilde{\varepsilon}^{-1} = \left(t_{x,y} \circ \pi\right)^{-1} = \pi^{-1} \circ \left(t_{x,y}\right)^{-1}$. Hence, (928) becomes

$$(\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ \pi \circ (\widetilde{\gamma} \circ \widetilde{\varepsilon})^{-1} = \underbrace{\widetilde{\varepsilon}^{-1}}_{=\pi^{-1} \circ \left(t_{x,y}\right)^{-1}} \circ t_{x,y} = \pi^{-1} \circ \underbrace{\left(t_{x,y}\right)^{-1} \circ t_{x,y}}_{=\mathrm{id}} = \pi^{-1}.$$

Thus, $\widetilde{\gamma} \circ \widetilde{\varepsilon}$ is an element of $\left\{\delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}}\right\}$ (by (926)) and satisfies $(\widetilde{\gamma} \circ \widetilde{\varepsilon}) \circ \pi \circ (\widetilde{\gamma} \circ \widetilde{\varepsilon})^{-1} = \pi^{-1}$. Hence, there exists a $\sigma \in \left\{\delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}}\right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$ (namely, $\sigma = \widetilde{\gamma} \circ \widetilde{\varepsilon}$). Hence, Claim 2 is proven in Case 2.

We have now proven that Claim 2 holds in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Claim 2 always holds.]

Thus, we have proven Claim 2. In other words, we have proven that Proposition 5.62 holds whenever $|X| = N$. This completes the induction step. The induction proof of Proposition 5.62 is thus complete. □

Proof of Theorem 5.61. If $X = \varnothing$, then Theorem 5.61 holds for simple reasons[439]. Hence, for the rest of this proof of Theorem 5.61, we can WLOG assume that we don't have $X = \varnothing$. Assume this.

There exists some $x \in X$ (since we don't have $X = \varnothing$). Consider this x. Hence, Proposition 5.62 shows that there exists a $\sigma \in \left\{\delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}}\right\}$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$. Consider this σ, and denote it by α. Thus, α is an element of $\left\{\delta^{(X \setminus \{x\} \to X)} \mid \delta \in S_{X \setminus \{x\}}\right\}$ and satisfies $\alpha \circ \pi \circ \alpha^{-1} = \pi^{-1}$.

[439] *Proof.* Assume that $X = \varnothing$. We must prove that Theorem 5.61 holds.

We have $X = \varnothing$. Thus, no $x \in X$ exists. Hence, $\left(\pi \circ \pi \circ \pi^{-1}\right)(x) = \pi^{-1}(x)$ holds for all $x \in X$ (indeed, this is vacuously true, since no $x \in X$ exists). In other words, we have $\pi \circ \pi \circ \pi^{-1} = \pi^{-1}$. Thus, there exists a $\sigma \in S_X$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$ (namely, $\sigma = \pi$). Hence, Theorem 5.61 holds.

We thus have proven that if $X = \varnothing$, then Theorem 5.61 holds. Qed.

We know that α is an element of $\left\{\delta^{(X\setminus\{x\}\to X)} \mid \delta \in S_{X\setminus\{x\}}\right\}$. In other words, there exists some $\delta \in S_{X\setminus\{x\}}$ such that $\alpha = \delta^{(X\setminus\{x\}\to X)}$. Consider this δ.

Now, Proposition 5.58 **(c)** (applied to $Y = X \setminus \{x\}$ and $\sigma = \delta$) yields that δ satisfies $\delta^{(X\setminus\{x\}\to X)} \in S_X$ and $\left(\delta^{-1}\right)^{(X\setminus\{x\}\to X)} = \left(\delta^{(X\setminus\{x\}\to X)}\right)^{-1}$. Now, $\alpha = \delta^{(X\setminus\{x\}\to X)} \in S_X$ and $\alpha \circ \pi \circ \alpha^{-1} = \pi^{-1}$. Hence, there exists a $\sigma \in S_X$ such that $\sigma \circ \pi \circ \sigma^{-1} = \pi^{-1}$ (namely, $\sigma = \alpha$). This proves Theorem 5.61. □

Solution to Exercise 5.22. We have now proven Proposition 5.56, Proposition 5.58, Proposition 5.59, Proposition 5.60, Proposition 5.62 and Theorem 5.61. Thus, Exercise 5.22 is solved. □

7.62. Solution to Exercise 5.23

We shall use the Iverson bracket notation introduced in Definition 3.48.

Lemma 7.130. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$. Let $j \in [n]$.
(a) Then, $\sum_{i\in[n]} [i \leq j] = j$.
(b) Let $\sigma \in S_n$. Then, $\sum_{i\in[n]} [\sigma(i) \leq \sigma(j)] = \sigma(j)$.

Proof of Lemma 7.130. **(a)** From $[n] = \{1, 2, \ldots, n\}$, we obtain

$$\sum_{i\in[n]} [i \leq j] = \underbrace{\sum_{i\in\{1,2,\ldots,n\}}}_{=\sum_{i=1}^{n}} [i \leq j] = \sum_{i=1}^{n} [i \leq j] = \sum_{i=1}^{j} \underbrace{[i \leq j]}_{\substack{=1 \\ \text{(since we have } i\leq j)}} + \sum_{i=j+1}^{n} \underbrace{[i \leq j]}_{\substack{=0 \\ \text{(since we don't have } i\leq j \\ \text{(because } i\geq j+1>j))}}$$
$$= \sum_{i=1}^{j} 1 + \underbrace{\sum_{i=j+1}^{n} 0}_{=0} = \sum_{i=1}^{j} 1 = j \cdot 1 = j.$$

This proves Lemma 7.130 **(a)**.

(b) Recall that S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$ (by the definition of S_n). In other words, S_n is the set of all permutations of the set $[n]$ (since $[n] = \{1, 2, \ldots, n\}$). Thus, σ is a permutation of the set $[n]$ (since $\sigma \in S_n$). In other words, σ is a bijection $[n] \to [n]$. Hence, we can substitute i for $\sigma(i)$ in the sum $\sum_{i\in[n]} [\sigma(i) \leq \sigma(j)]$. We thus obtain

$$\sum_{i\in[n]} [\sigma(i) \leq \sigma(j)] = \sum_{i\in[n]} [i \leq \sigma(j)] = \sigma(j)$$

(by Lemma 7.130 **(a)**, applied to $\sigma(j)$ instead of j). This proves Lemma 7.130 **(b)**. □

Lemma 7.131. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$. Let $\sigma \in S_n$. Let $a_1, a_2, \ldots, a_n$ be any n numbers. (Here, "number" means "real number" or "complex number" or "rational number", as you prefer; this makes no difference.) Then,

$$\sum_{\substack{1 \leq i < j \leq n; \\ \sigma(i) > \sigma(j)}} \left(a_j - a_i\right) = \sum_{i \in [n]} \sum_{j \in [n]} [i < j] [\sigma(i) > \sigma(j)] \left(a_j - a_i\right).$$

[Here, the summation sign "$\sum_{\substack{1 \leq i < j \leq n; \\ \sigma(i) > \sigma(j)}}$" means "$\sum_{\substack{(i,j) \in \{1,2,\ldots,n\}^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}}$"; this is a sum over all inversions of σ.]

Proof of Lemma 7.131. If $\mathcal{A}$ and $\mathcal{B}$ are two logical statements, then $[\mathcal{A} \wedge \mathcal{B}] = [\mathcal{A}] [\mathcal{B}]$ (by Exercise 3.12 **(b)**). Hence, if $\mathcal{A}$ and $\mathcal{B}$ are two logical statements, then

$$[\mathcal{A}] [\mathcal{B}] = [\mathcal{A} \wedge \mathcal{B}] = [\mathcal{A} \text{ and } \mathcal{B}] \tag{929}$$

(since "$\mathcal{A} \wedge \mathcal{B}$" means "$\mathcal{A}$ and $\mathcal{B}$").

Recall that the summation sign "$\sum_{\substack{1 \leq i < j \leq n; \\ \sigma(i) > \sigma(j)}}$" means "$\sum_{\substack{(i,j) \in \{1,2,\ldots,n\}^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}}$". Hence,

$$\sum_{\substack{1 \leq i < j \leq n; \\ \sigma(i) > \sigma(j)}} \left(a_j - a_i\right) = \sum_{\substack{(i,j) \in \{1,2,\ldots,n\}^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}} \left(a_j - a_i\right) = \sum_{\substack{(i,j) \in [n]^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}} \left(a_j - a_i\right)$$

(since $\{1, 2, \ldots, n\} = [n]$). Comparing this with

$$\begin{aligned}
&\underbrace{\sum_{i \in [n]} \sum_{j \in [n]}}_{= \sum_{(i,j) \in [n]^2}} \underbrace{[i < j] [\sigma(i) > \sigma(j)]}_{\substack{= [i<j \text{ and } \sigma(i) > \sigma(j)] \\ \text{(by (929) (applied} \\ \text{to } \mathcal{A} = (i<j) \text{ and } \mathcal{B} = (\sigma(i) > \sigma(j))))}} \left(a_j - a_i\right) \\
&= \sum_{(i,j) \in [n]^2} [i < j \text{ and } \sigma(i) > \sigma(j)] \left(a_j - a_i\right) \\
&= \sum_{\substack{(i,j) \in [n]^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}} \underbrace{[i < j \text{ and } \sigma(i) > \sigma(j)]}_{\substack{=1 \\ \text{(since } i<j \text{ and } \sigma(i) > \sigma(j))}} \left(a_j - a_i\right) \\
&\qquad + \sum_{\substack{(i,j) \in [n]^2; \\ \text{not } (i<j \text{ and } \sigma(i) > \sigma(j))}} \underbrace{[i < j \text{ and } \sigma(i) > \sigma(j)]}_{\substack{=0 \\ \text{(since we don't have } (i<j \text{ and } \sigma(i) > \sigma(j)))}} \left(a_j - a_i\right) \\
&= \sum_{\substack{(i,j) \in [n]^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}} \left(a_j - a_i\right) + \underbrace{\sum_{\substack{(i,j) \in [n]^2; \\ \text{not } (i<j \text{ and } \sigma(i) > \sigma(j))}} 0 \left(a_j - a_i\right)}_{=0} = \sum_{\substack{(i,j) \in [n]^2; \\ i<j \text{ and } \sigma(i) > \sigma(j)}} \left(a_j - a_i\right),
\end{aligned}$$

we obtain

$$\sum_{\substack{1\le i<j\le n;\\ \sigma(i)>\sigma(j)}} \left(a_j - a_i\right) = \sum_{i\in[n]} \sum_{j\in[n]} [i<j] [\sigma(i) > \sigma(j)] \left(a_j - a_i\right).$$

This proves Lemma 7.131. $\square$

Lemma 7.132. Let $n \in \mathbb{N}$. Let $[n] = \{1, 2, \ldots, n\}$. Let $\sigma \in S_n$. Let $i \in [n]$ and $j \in [n]$. Then,

$$[i<j] [\sigma(i) > \sigma(j)] - [j<i] [\sigma(j) > \sigma(i)] = [i \le j] - [\sigma(i) \le \sigma(j)].$$

Proof of Lemma 7.132. If $i = j$, then all four truth values $[i<j]$, $[\sigma(i) > \sigma(j)]$, $[j<i]$ and $[\sigma(j) > \sigma(i)]$ equal 0 (because $i = j$ and $\sigma(i) = \sigma(j)$), whereas the two truth values $[i \le j]$ and $[\sigma(i) \le \sigma(j)]$ equal 1 (since $i = j \le j$ and $\sigma(i) = \sigma(j) \le \sigma(j)$). Hence, if $i = j$, then Lemma 7.132 boils down to the equality $0 \cdot 0 - 0 \cdot 0 = 1 - 1$, which is obvious. Thus, for the rest of the proof of Lemma 7.132, we WLOG assume that $i \ne j$.

The map σ is a permutation of $[n]$ (since $\sigma \in S_n$), and thus is bijective. Hence, σ is injective. Thus, from $i \ne j$, we conclude that $\sigma(i) \ne \sigma(j)$. Therefore, $\sigma(i) \le \sigma(j)$ holds if and only if $\sigma(i) < \sigma(j)$. We thus have the following equivalence of statements:

$$(\sigma(i) \le \sigma(j)) \iff (\sigma(i) < \sigma(j)) \iff (\sigma(j) > \sigma(i)).$$

Therefore, $[\sigma(i) \le \sigma(j)] = [\sigma(j) > \sigma(i)]$.

Also, $i \le j$ holds if and only if $i < j$ (since $i \ne j$). Thus, the statements $(i \le j)$ and $(i < j)$ are equivalent. Hence, $[i \le j] = [i < j]$.

But any two statements $\mathcal{A}$ and $\mathcal{B}$ satisfy

$$[\mathcal{A}] - [\mathcal{B}] = [\mathcal{A}] [\text{not } \mathcal{B}] - [\text{not } \mathcal{A}] [\mathcal{B}]. \tag{930}$$

(Indeed, if $\mathcal{A}$ and $\mathcal{B}$ are two statements, then

$$[\mathcal{A}] \underbrace{[\text{not } \mathcal{B}]}_{\substack{=1-[\mathcal{B}] \\ \text{(by Exercise 3.12 \textbf{(b)} (applied to } \mathcal{B} \\ \text{instead of } \mathcal{A}))}} - \underbrace{[\text{not } \mathcal{A}]}_{\substack{=1-[\mathcal{A}] \\ \text{(by Exercise 3.12 \textbf{(b)})}}} [\mathcal{B}]$$
$$= [\mathcal{A}] (1 - [\mathcal{B}]) - (1 - [\mathcal{A}]) [\mathcal{B}] = [\mathcal{A}] - [\mathcal{A}] [\mathcal{B}] - [\mathcal{B}] + [\mathcal{A}] [\mathcal{B}] = [\mathcal{A}] - [\mathcal{B}].$$

)

Applying (930) to $\mathcal{A} = (i \le j)$ and $\mathcal{B} = (\sigma(i) \le \sigma(j))$, we obtain

$$\begin{aligned} & [i \le j] - [\sigma(i) \le \sigma(j)] \\ &= \underbrace{[i \le j]}_{=[i<j]} \underbrace{[\text{not } \sigma(i) \le \sigma(j)]}_{\substack{=[\sigma(i)>\sigma(j)] \\ \text{(since (not } \sigma(i)\le\sigma(j)) \text{ is} \\ \text{equivalent to } (\sigma(i)>\sigma(j)))}} - \underbrace{[\text{not } i \le j]}_{\substack{=[j<i] \\ \text{(since (not } i\le j) \text{ is} \\ \text{equivalent to } (j<i))}} \underbrace{[\sigma(i) \le \sigma(j)]}_{=[\sigma(j)>\sigma(i)]} \\ &= [i<j] [\sigma(i) > \sigma(j)] - [j<i] [\sigma(j) > \sigma(i)]. \end{aligned}$$

This proves Lemma 7.132. □

Solving Exercise 5.23 is now a mere matter of computation:

Solution to Exercise 5.23. Let $[n] = \{1, 2, \ldots, n\}$. Lemma 7.131 yields

$$\begin{aligned}
&\sum_{\substack{1\leq i<j\leq n;\\ \sigma(i)>\sigma(j)}} \left(a_j - a_i\right) \\
&= \sum_{i\in[n]} \sum_{j\in[n]} \underbrace{[i<j]\,[\sigma(i)>\sigma(j)]\,\left(a_j - a_i\right)}_{=[i<j][\sigma(i)>\sigma(j)]a_j - [i<j][\sigma(i)>\sigma(j)]a_i} \\
&= \sum_{i\in[n]} \sum_{j\in[n]} \left([i<j]\,[\sigma(i)>\sigma(j)]\,a_j - [i<j]\,[\sigma(i)>\sigma(j)]\,a_i\right) \\
&= \sum_{i\in[n]} \sum_{j\in[n]} [i<j]\,[\sigma(i)>\sigma(j)]\,a_j - \sum_{i\in[n]} \sum_{j\in[n]} [i<j]\,[\sigma(i)>\sigma(j)]\,a_i. \qquad (931)
\end{aligned}$$

But

$$\begin{aligned}
&\underbrace{\sum_{i\in[n]} \sum_{j\in[n]}}_{=\sum_{j\in[n]}\sum_{i\in[n]}} [i<j]\,[\sigma(i)>\sigma(j)]\,a_i \\
&= \sum_{j\in[n]} \sum_{i\in[n]} [i<j]\,[\sigma(i)>\sigma(j)]\,a_i = \sum_{i\in[n]} \sum_{j\in[n]} [j<i]\,[\sigma(j)>\sigma(i)]\,a_j
\end{aligned}$$

(here, we have renamed the summation indices j and i as i and j, respectively).

Hence, (931) becomes

$$\begin{aligned}
&\sum_{\substack{1\leq i<j\leq n;\\ \sigma(i)>\sigma(j)}} \left(a_j - a_i\right) \\
&= \sum_{i\in[n]} \sum_{j\in[n]} [i<j]\,[\sigma(i)>\sigma(j)]\,a_j - \underbrace{\sum_{i\in[n]} \sum_{j\in[n]} [i<j]\,[\sigma(i)>\sigma(j)]\,a_i}_{=\sum_{i\in[n]}\sum_{j\in[n]}[j<i][\sigma(j)>\sigma(i)]a_j} \\
&= \sum_{i\in[n]} \sum_{j\in[n]} [i<j]\,[\sigma(i)>\sigma(j)]\,a_j - \sum_{i\in[n]} \sum_{j\in[n]} [j<i]\,[\sigma(j)>\sigma(i)]\,a_j \\
&= \underbrace{\sum_{i\in[n]} \sum_{j\in[n]}}_{=\sum_{j\in[n]}\sum_{i\in[n]}} \underbrace{\left([i<j]\,[\sigma(i)>\sigma(j)]\,a_j - [j<i]\,[\sigma(j)>\sigma(i)]\,a_j\right)}_{=([i<j][\sigma(i)>\sigma(j)]-[j<i][\sigma(j)>\sigma(i)])a_j} \\
&= \sum_{j\in[n]} \sum_{i\in[n]} \underbrace{\left([i<j]\,[\sigma(i)>\sigma(j)] - [j<i]\,[\sigma(j)>\sigma(i)]\right)}_{\substack{=[i\leq j]-[\sigma(i)\leq\sigma(j)]\\ \text{(by Lemma 7.132)}}} a_j \\
&= \sum_{j\in[n]} \sum_{i\in[n]} \left([i\leq j] - [\sigma(i)\leq\sigma(j)]\right) a_j = \sum_{j\in[n]} \underbrace{\left(\sum_{i\in[n]} \left([i\leq j] - [\sigma(i)\leq\sigma(j)]\right)\right)}_{=\sum_{i\in[n]}[i\leq j]-\sum_{i\in[n]}[\sigma(i)\leq\sigma(j)]} a_j \\
&= \sum_{j\in[n]} \left(\underbrace{\sum_{i\in[n]} [i\leq j]}_{\substack{=j\\ \text{(by Lemma 7.130 \textbf{(a)})}}} - \underbrace{\sum_{i\in[n]} [\sigma(i)\leq\sigma(j)]}_{\substack{=\sigma(j)\\ \text{(by Lemma 7.130 \textbf{(b)})}}} \right) a_j \\
&= \underbrace{\sum_{j\in[n]}}_{=\sum_{j=1}^{n}} \underbrace{(j-\sigma(j))\,a_j}_{=a_j(j-\sigma(j))} = \sum_{j=1}^{n} a_j\,(j-\sigma(j)) = \sum_{i=1}^{n} a_i\,(i-\sigma(i))
\end{aligned}$$

(here, we have renamed the summation index j as i). This solves Exercise 5.23. □

7.63. Solution to Exercise 5.24

Solution to Exercise 5.24. We have $\pi \in S_n$. In other words, π is a permutation of $\{1,2,\ldots,n\}$ (since S_n is the set of all permutations of $\{1,2,\ldots,n\}$). In other words, π is a bijection $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$. Hence, we can substitute $\pi(i)$ for i in the sum $\sum_{i\in\{1,2,\ldots,n\}} i^2$. We thus obtain

$$\sum_{i\in\{1,2,\ldots,n\}} i^2 = \sum_{i\in\{1,2,\ldots,n\}} (\pi(i))^2 . \tag{932}$$

(a) Exercise 5.23 (applied to $\sigma = \pi$ and $a_k = \pi(k) + k$) yields

$$\begin{aligned}
\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} ((\pi(j)+j) - (\pi(i)+i)) &= \underbrace{\sum_{i=1}^{n}}_{=\sum_{i\in\{1,2,\ldots,n\}}} \underbrace{(\pi(i)+i)(i-\pi(i))}_{\substack{=(i+\pi(i))(i-\pi(i))\\ =i^2-(\pi(i))^2\\ \text{(since } (x+y)(x-y)=x^2-y^2\\ \text{for any two numbers } x \text{ and } y)}} \\
&= \sum_{i\in\{1,2,\ldots,n\}} \left(i^2 - (\pi(i))^2\right) \\
&= \underbrace{\sum_{i\in\{1,2,\ldots,n\}} i^2}_{\substack{=\sum_{i\in\{1,2,\ldots,n\}} (\pi(i))^2\\ \text{(by (932))}}} - \sum_{i\in\{1,2,\ldots,n\}} (\pi(i))^2 \\
&= \sum_{i\in\{1,2,\ldots,n\}} (\pi(i))^2 - \sum_{i\in\{1,2,\ldots,n\}} (\pi(i))^2 = 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
0 &= \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} \underbrace{((\pi(j)+j) - (\pi(i)+i))}_{=(\pi(j)-\pi(i))-(i-j)} \\
&= \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} ((\pi(j)-\pi(i)) - (i-j)) \\
&= \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} (\pi(j)-\pi(i)) - \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} (i-j) .
\end{aligned}$$

Adding $\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} (i-j)$ to both sides of this equality, we obtain

$$\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} (i-j) = \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)>\pi(j)}} (\pi(j)-\pi(i)) .$$

This solves Exercise 5.24 **(a)**.

(b) Let w_0 denote the permutation in S_n which sends each $k \in \{1,2,\ldots,n\}$ to $n+1-k$. Define a permutation $\sigma \in S_n$ by $\sigma = w_0 \circ \pi$. Thus, each $k \in \{1,2,\ldots,n\}$ satisfies

$$\underbrace{\sigma}_{=w_0\circ\pi}(k) = (w_0\circ\pi)(k) = w_0(\pi(k)) = n+1-\pi(k) \tag{933}$$

(by the definition of w_0).

For any $(i,j) \in \{1,2,\ldots,n\}^2$, we have the following chain of logical equivalences:

$$\left(\underbrace{\sigma(i)}_{\substack{=n+1-\pi(i) \\ \text{(by (933)} \\ \text{(applied to } k=i))}} > \underbrace{\sigma(j)}_{\substack{=n+1-\pi(j) \\ \text{(by (933)} \\ \text{(applied to } k=j))}} \right) \iff (n+1-\pi(i) > n+1-\pi(j)) \iff (\pi(i) < \pi(j)).$$

Thus, for any $(i,j) \in \{1,2,\ldots,n\}^2$, the condition $(\sigma(i) > \sigma(j))$ is equivalent to $(\pi(i) < \pi(j))$. Hence, the summation sign "$\sum_{\substack{1\le i<j\le n; \\ \sigma(i)>\sigma(j)}}$" can be rewritten as "$\sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}}$".

In other words, we have

$$\sum_{\substack{1\le i<j\le n; \\ \sigma(i)>\sigma(j)}} = \sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}}$$

(an equality between summation signs). Now, Exercise 5.24 **(a)** (applied to σ instead of π) yields

$$\sum_{\substack{1\le i<j\le n; \\ \sigma(i)>\sigma(j)}} (\sigma(j) - \sigma(i)) = \underbrace{\sum_{\substack{1\le i<j\le n; \\ \sigma(i)>\sigma(j)}}}_{= \sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}}} \underbrace{(i-j)}_{=-(j-i)} = \sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}} (-(j-i)) = -\sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}} (j-i).$$

Comparing this with

$$\begin{aligned}
&\underbrace{\sum_{\substack{1\le i<j\le n; \\ \sigma(i)>\sigma(j)}}}_{= \sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}}} \left(\underbrace{\sigma(j)}_{\substack{=n+1-\pi(j) \\ \text{(by (933)} \\ \text{(applied to } k=j))}} - \underbrace{\sigma(i)}_{\substack{=n+1-\pi(i) \\ \text{(by (933)} \\ \text{(applied to } k=i))}} \right) \\
&= \sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}} \underbrace{((n+1-\pi(j)) - (n+1-\pi(i)))}_{=-(\pi(j)-\pi(i))} = \sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}} (-(\pi(j)-\pi(i))) \\
&= -\sum_{\substack{1\le i<j\le n; \\ \pi(i)<\pi(j)}} (\pi(j)-\pi(i)),
\end{aligned}$$

we obtain

$$-\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}} (\pi(j)-\pi(i)) = -\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}} (j-i).$$

Thus,

$$\sum_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}} (\pi(j)-\pi(i)) = \sum_{\substack{1\leq i<j\leq n;\\ \pi(i)<\pi(j)}} (j-i).$$

This solves Exercise 5.24 **(b)**. □

7.64. Solution to Exercise 5.25

Throughout Section 7.64, we shall use the notations introduced in Definition 7.74 and in Definition 5.30.

Before we step to the solution of Exercise 5.25, let us prove a simple lemma:

Lemma 7.133. Let $n \in \mathbb{N}$. Let $\pi \in S_n$. Let u and v be two elements of $[n]$. Then,

$$t_{\pi(u),\pi(v)} \circ \pi = \pi \circ t_{u,v}.$$

Proof of Lemma 7.133. We have $[n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$).

Thus, S_n is the set of all permutations of $[n]$. Hence, all three maps $t_{\pi(u),\pi(v)}$, π and $t_{u,v}$ are permutations of $[n]$, that is, bijections from $[n]$ to $[n]$. Thus, the map π is a bijection, and hence is injective.

Fix $k \in [n]$. We shall prove that $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$.

We are in one of the following three cases:

Case 1: We have $k = u$.

Case 2: We have $k = v$.

Case 3: We have neither $k = u$ nor $k = v$.

Let us first consider Case 1. In this case, we have $k = u$. But Lemma 7.75 **(a)** (applied to $i = u$ and $j = v$) yields $t_{u,v}(u) = v$. Also,

$$\left(t_{\pi(u),\pi(v)} \circ \pi\right)\left(\underbrace{k}_{=u}\right) = \left(t_{\pi(u),\pi(v)} \circ \pi\right)(u) = t_{\pi(u),\pi(v)}(\pi(u)) = \pi(v)$$

(by Lemma 7.75 **(a)** (applied to $i = \pi(u)$ and $j = \pi(v)$)). Comparing this with

$$(\pi \circ t_{u,v})\left(\underbrace{k}_{=u}\right) = (\pi \circ t_{u,v})(u) = \pi\left(\underbrace{t_{u,v}(u)}_{=v}\right) = \pi(v),$$

we obtain $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$. Hence, $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$ is proven in Case 1.

The argument in Case 2 is analogous, and we leave it to the reader.

Let us now consider Case 3. In this case, we have neither $k = u$ nor $k = v$. Thus, $k \in [n] \setminus \{u, v\}$. Hence, Lemma 7.75 **(c)** (applied to $i = u$ and $j = v$) yields $t_{u,v}(k) = k$. Also, recall that we have neither $k = u$ nor $k = v$. Thus, we have neither $\pi(k) = \pi(u)$ nor $\pi(k) = \pi(v)$ (since the map π is injective). In other words, $\pi(k) \in [n] \setminus \{\pi(u), \pi(v)\}$. Thus, Lemma 7.75 **(c)** (applied to $\pi(u)$, $\pi(v)$ and $\pi(k)$ instead of i, j and k) yields $t_{\pi(u),\pi(v)}(\pi(k)) = \pi(k)$. Now,

$$\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = t_{\pi(u),\pi(v)}(\pi(k)) = \pi(k).$$

Comparing this with

$$(\pi \circ t_{u,v})(k) = \pi\left(\underbrace{t_{u,v}(k)}_{=k}\right) = \pi(k),$$

we obtain $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$. Hence, $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$ is proven in Case 3.

We have now proven $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$ in each of the three Cases 1, 2 and 3. Thus, $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$ always holds.

Forget now that we fixed k. We thus have shown that $\left(t_{\pi(u),\pi(v)} \circ \pi\right)(k) = (\pi \circ t_{u,v})(k)$ for each $k \in [n]$. In other words, $t_{\pi(u),\pi(v)} \circ \pi = \pi \circ t_{u,v}$ (since both $t_{\pi(u),\pi(v)} \circ \pi$ and $\pi \circ t_{u,v}$ are maps from $[n]$ to $[n]$). Thus, Lemma 7.133 is proven. □

We can now solve Exercise 5.25:

Solution to Exercise 5.25. Recall that $(i_1, i_2, \ldots, i_n) \in [1] \times [2] \times \cdots \times [n]$. Thus,

$$i_j \in [j] \qquad \text{for each } j \in [n]. \tag{934}$$

The definition of σ_0 shows that

$$\sigma_0 = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{0,i_0} = (\text{a composition of 0 permutations}) = \operatorname{id}. \tag{935}$$

The definition of σ_n shows that

$$\sigma_n = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n} = \sigma \tag{936}$$

(since $\sigma = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{n,i_n}$).

For each $k \in [n]$, we have

$$\sigma_k = \sigma_{k-1} \circ t_{k,i_k}. \tag{937}$$

[*Proof of (937):* Let $k \in [n]$. The definition of σ_{k-1} yields $\sigma_{k-1} = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k-1,i_{k-1}}$. But the definition of σ_k yields

$$\sigma_k = t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k,i_k} = \underbrace{\left(t_{1,i_1} \circ t_{2,i_2} \circ \cdots \circ t_{k-1,i_{k-1}}\right)}_{=\sigma_{k-1}} \circ t_{k,i_k} = \sigma_{k-1} \circ t_{k,i_k}.$$

This proves (937).]

(a) Let $i \in [n]$. We shall prove that

$$\sigma_k (i) = i \qquad \text{for each } k \in \{0, 1, \ldots, i-1\}. \tag{938}$$

[*Proof of (938):* We shall prove (938) by induction on k:

Induction base: From (935), we obtain $\sigma_0 = \text{id}$. Hence, $\sigma_0 (i) = \text{id} (i) = i$. In other words, (938) holds for $k = 0$. This completes the induction base.

Induction step: Let $p \in \{1, 2, \ldots, i-1\}$. Assume that (938) holds for $k = p - 1$. We must prove that (938) holds for $k = p$.

We assumed that (938) holds for $k = p - 1$. In other words, $\sigma_{p-1} (i) = i$. Also, (937) (applied to $k = p$) yields $\sigma_p = \sigma_{p-1} \circ t_{p,i_p}$ (since $p \in [n]$).

We have $p \in \{1, 2, \ldots, i-1\}$, thus $p \leq i - 1 < i$ and therefore $p \neq i$. In other words, $i \neq p$. Also, (934) (applied to $j = p$) yields $i_p \in [p] = \{1, 2, \ldots, p\}$; thus, $i_p \leq p < i$ and therefore $i_p \neq i$. In other words, $i \neq i_p$. So we know that $i \neq p$ and $i \neq i_p$.

Hence, $i \in [n] \setminus \{p, i_p\}$. Thus, Lemma 7.75 **(c)** (applied to p, i_p and i instead of i, j and k) yields $t_{p,i_p} (i) = i$.

Now,

$$\underbrace{\sigma_p}_{=\sigma_{p-1} \circ t_{p,i_p}} (i) = \left(\sigma_{p-1} \circ t_{p,i_p}\right) (i) = \sigma_{p-1} \left(\underbrace{t_{p,i_p} (i)}_{=i} \right) = \sigma_{p-1} (i) = i.$$

In other words, (938) holds for $k = p$. This completes the induction step. Thus, (938) is proven by induction.]

Thus, we have shown that $\sigma_k (i) = i$ for each $k \in \{0, 1, \ldots, i-1\}$. This solves Exercise 5.25 **(a)**.

(b) Let $k \in [n]$. The definition of m_k yields $m_k = \sigma_k (k)$.

Assume (for the sake of contradiction) that $m_k \notin [k]$. Combining $m_k = \sigma_k (k) \in [n]$ with $m_k \notin [k]$, we obtain

$$m_k \in \underbrace{[n]}_{=\{1,2,\ldots,n\}} \setminus \underbrace{[k]}_{=\{1,2,\ldots,k\}} = \{1, 2, \ldots, n\} \setminus \{1, 2, \ldots, k\} = \{k+1, k+2, \ldots, n\}.$$

Thus, $m_k \geq k + 1$, so that $k \leq m_k - 1$. Hence, $k \in \{0, 1, \ldots, m_k - 1\}$ (because $k \in \mathbb{N}$). Hence, Exercise 5.25 **(a)** (applied to $i = m_k$) yields $\sigma_k (m_k) = m_k = \sigma_k (k)$. Since σ_k is injective (because $\sigma_k \in S_n$), this yields $m_k = k \leq m_k - 1 < m_k$. But this is clearly absurd. This contradiction shows that our assumption (that $m_k \notin [k]$) was false. Hence, we must have $m_k \in [k]$. This solves Exercise 5.25 **(b)**.

(c) Let $k \in [n]$.

Thus, $k \geq 1$, and therefore $k - 1 \in \mathbb{N}$, so that $k - 1 \in \{0, 1, \ldots, k-1\}$.

Hence, Exercise 5.25 **(a)** (applied to k and $k - 1$ instead of i and k) yields

$$\sigma_{k-1} (k) = k. \tag{939}$$

On the other hand, Lemma 7.75 **(b)** (applied to k and i_k instead of i and j) yields $t_{k,i_k}(i_k) = k$. Furthermore, Lemma 7.75 **(a)** (applied to k and i_k instead of i and j) yields $t_{k,i_k}(k) = i_k$.

But (937) yields $\sigma_k = \sigma_{k-1} \circ t_{k,i_k}$. Hence,

$$\underbrace{\sigma_k}_{=\sigma_{k-1}\circ t_{k,i_k}} (i_k) = \left(\sigma_{k-1} \circ t_{k,i_k}\right)(i_k) = \sigma_{k-1}\left(\underbrace{t_{k,i_k}(i_k)}_{=k}\right) = \sigma_{k-1}(k) = k.$$

This solves Exercise 5.25 **(c)**.

(d) Let $k \in [n]$. Then, (937) yields $\sigma_k = \sigma_{k-1} \circ t_{k,i_k}$. In our above solution to Exercise 5.25 **(c)**, we have shown that $\sigma_{k-1}(k) = k$ and $t_{k,i_k}(k) = i_k$. Also, the definition of m_k yields

$$m_k = \underbrace{\sigma_k}_{=\sigma_{k-1}\circ t_{k,i_k}} (k) = \left(\sigma_{k-1} \circ t_{k,i_k}\right)(k) = \sigma_{k-1}\left(\underbrace{t_{k,i_k}(k)}_{=i_k}\right) = \sigma_{k-1}(i_k),$$

so that

$$\sigma_{k-1}(i_k) = m_k. \qquad (940)$$

Define $\pi \in S_n$ by $\pi = \sigma_{k-1}$. Then, $\underbrace{\pi}_{=\sigma_{k-1}}(k) = \sigma_{k-1}(k) = k$ and $\underbrace{\pi}_{=\sigma_{k-1}}(i_k) = \sigma_{k-1}(i_k) = m_k$.

Now, Lemma 7.133 (applied to $u = k$ and $v = i_k$) yields

$$t_{\pi(k),\pi(i_k)} \circ \pi = \underbrace{\pi}_{=\sigma_{k-1}} \circ t_{k,i_k} = \sigma_{k-1} \circ t_{k,i_k} = \sigma_k$$

(since $\sigma_k = \sigma_{k-1} \circ t_{k,i_k}$). In view of $\pi(k) = k$ and $\pi(i_k) = m_k$, this rewrites as $t_{k,m_k} \circ \pi = \sigma_k$. Hence, $\sigma_k = t_{k,m_k} \circ \underbrace{\pi}_{=\sigma_{k-1}} = t_{k,m_k} \circ \sigma_{k-1}$. This solves Exercise 5.25 **(d)**.

(e) We shall show that

$$\sigma_p^{-1} = t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{p,m_p} \qquad \text{for each } p \in \{0, 1, \ldots, n\}. \qquad (941)$$

[*Proof of (941):* We shall prove (941) by induction on p:

Induction base: We have $\sigma_0 = \text{id}$ (by (935)) and thus $\sigma_0^{-1} = \text{id}^{-1} = \text{id}$. Comparing this with

$$t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{0,m_0} = (\text{a composition of } 0 \text{ permutations}) = \text{id},$$

we obtain $\sigma_0^{-1} = t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{0,m_0}$. In other words, (941) holds for $p = 0$. This completes the induction base.

Induction step: Let $k \in \{1, 2, \ldots, n\}$. Assume that (941) holds for $p = k - 1$. We must prove that (941) holds for $p = k$.

We have assumed that (941) holds for $p = k-1$. That is, we have

$$\sigma_{k-1}^{-1} = t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{k-1,m_{k-1}}.$$

Lemma 7.75 **(d)** (applied to k and m_k instead of i and j) yields $t_{k,m_k} \circ t_{k,m_k} = \mathrm{id}$. Thus, $t_{k,m_k}^{-1} = t_{k,m_k}$.

But $k \in \{1, 2, \ldots, n\} = [n]$. Hence, Exercise 5.25 **(d)** yields $\sigma_k = t_{k,m_k} \circ \sigma_{k-1}$. Thus,

$$\sigma_k^{-1} = \left(t_{k,m_k} \circ \sigma_{k-1}\right)^{-1} = \underbrace{\sigma_{k-1}^{-1}}_{=t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{k-1,m_{k-1}}} \circ \underbrace{t_{k,m_k}^{-1}}_{=t_{k,m_k}}$$
$$= \left(t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{k-1,m_{k-1}}\right) \circ t_{k,m_k} = t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{k,m_k}.$$

In other words, (941) holds for $p = k$. This completes the induction step. Thus, (941) is proven by induction.]

Now, $n \in \{0, 1, \ldots, n\}$. Hence, (941) (applied to $p = n$) yields $\sigma_n^{-1} = t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{n,m_n}$. In view of (936), this rewrites as $\sigma^{-1} = t_{1,m_1} \circ t_{2,m_2} \circ \cdots \circ t_{n,m_n}$. This solves Exercise 5.25 **(e)**.

(f) For each permutation $\tau \in S_n$, we define a number $z(\tau)$ by

$$z(\tau) = \sum_{k=1}^{n} x_k y_{\tau(k)}.$$

We shall show that

$$z(\sigma_{p-1}) - z(\sigma_p) = \left(x_{i_p} - x_p\right)\left(y_{m_p} - y_p\right) \qquad \text{for each } p \in [n]. \tag{942}$$

[*Proof of (942):* Let $p \in [n]$. Applying (937) to $k = p$, we obtain $\sigma_p = \sigma_{p-1} \circ t_{p,i_p}$. Hence, if $p = i_p$, then (942) holds[440]. Thus, for the rest of this proof of (942), we WLOG assume that $p \neq i_p$. Hence, t_{p,i_p} is an actual transposition (not the identity map).

From $\sigma_p = \sigma_{p-1} \circ t_{p,i_p}$, we obtain

$$\sigma_p(p) = \left(\sigma_{p-1} \circ t_{p,i_p}\right)(p) = \sigma_{p-1}\left(\underbrace{t_{p,i_p}(p)}_{=i_p}\right) = \sigma_{p-1}\left(i_p\right),$$

so that

$$\sigma_{p-1}\left(i_p\right) = \sigma_p(p) = m_p \tag{943}$$

[440] *Proof.* Assume that $p = i_p$. Thus, $i_p = p$, so that $x_{i_p} - x_p = x_p - x_p = 0$. Hence, the right hand side of (942) equals 0. Also, $\sigma_p = \sigma_{p-1} \circ \underbrace{t_{p,i_p}}_{\substack{=\mathrm{id} \\ (\text{since } p = i_p)}} = \sigma_{p-1}$, so that $z(\sigma_{p-1}) - z(\sigma_p) = z(\sigma_{p-1}) - z(\sigma_{p-1}) = 0$. Thus, the left hand side of (942) equals 0 as well. Hence, the equality (942) holds (since both its right hand side and its left hand side equal 0).

(since the definition of m_p yields $m_p = \sigma_p(p)$).

From $\sigma_p = \sigma_{p-1} \circ t_{p,i_p}$, we also obtain

$$\sigma_p\left(i_p\right) = \left(\sigma_{p-1} \circ t_{p,i_p}\right)\left(i_p\right) = \sigma_{p-1}\left(\underbrace{t_{p,i_p}\left(i_p\right)}_{=p}\right) = \sigma_{p-1}(p),$$

so that

$$\sigma_{p-1}(p) = \sigma_p\left(i_p\right) = p \tag{944}$$

(by Exercise 5.25 **(c)**, applied to $k = p$).

Every $k \in [n]$ satisfying $k \neq p$ and $k \neq i_p$ satisfies

$$\sigma_{p-1}(k) = \sigma_p(k) \tag{945}$$

[441]. Now, the definition of $z\left(\sigma_{p-1}\right)$ yields

$$\begin{aligned} z\left(\sigma_{p-1}\right) &= \sum_{k=1}^{n} x_k y_{\sigma_{p-1}(k)} = x_p \underbrace{y_{\sigma_{p-1}(p)}}_{\substack{=y_p \\ \text{(by (944))}}} + x_{i_p} \underbrace{y_{\sigma_{p-1}\left(i_p\right)}}_{\substack{=y_{m_p} \\ \text{(by (943))}}} + \sum_{\substack{k\in[n];\\ k\neq p \text{ and } k\neq i_p}} x_k \underbrace{y_{\sigma_{p-1}(k)}}_{\substack{=y_{\sigma_p(k)} \\ \text{(by (945))}}} \\ &\qquad \left(\begin{array}{c}\text{here, we have split the addends for } k=p \text{ and} \\ \text{for } k=i_p \text{ from the sum (and these are} \\ \text{two distinct addends, since } p\neq i_p\text{)}\end{array}\right) \\ &= x_p y_p + x_{i_p} y_{m_p} + \sum_{\substack{k\in[n];\\ k\neq p \text{ and } k\neq i_p}} x_k y_{\sigma_p(k)}. \end{aligned}$$

On the other hand, the definition of $z\left(\sigma_p\right)$ yields

$$\begin{aligned} z\left(\sigma_p\right) &= \sum_{k=1}^{n} x_k y_{\sigma_p(k)} = x_p \underbrace{y_{\sigma_p(p)}}_{\substack{=y_{m_p} \\ \text{(since } \sigma_p(p)=m_p\text{)}}} + x_{i_p} \underbrace{y_{\sigma_p\left(i_p\right)}}_{\substack{=y_p \\ \text{(since } \sigma_p\left(i_p\right)=p\text{)}}} + \sum_{\substack{k\in[n];\\ k\neq p \text{ and } k\neq i_p}} x_k y_{\sigma_p(k)} \\ &\qquad \left(\begin{array}{c}\text{here, we have split the addends for } k=p \text{ and} \\ \text{for } k=i_p \text{ from the sum (and these are} \\ \text{two distinct addends, since } p\neq i_p\text{)}\end{array}\right) \\ &= x_p y_{m_p} + x_{i_p} y_p + \sum_{\substack{k\in[n];\\ k\neq p \text{ and } k\neq i_p}} x_k y_{\sigma_p(k)}. \end{aligned}$$

[441] *Proof:* Let $k \in [n]$ be such that $k \neq p$ and $k \neq i_p$. Thus, $t_{p,i_p}(k) = k$. But $\sigma_p = \sigma_{p-1} \circ t_{p,i_p}$; hence, $\sigma_p(k) = \left(\sigma_{p-1} \circ t_{p,i_p}\right)(k) = \sigma_{p-1}\left(\underbrace{t_{p,i_p}(k)}_{=k}\right) = \sigma_{p-1}(k)$, so that $\sigma_{p-1}(k) = \sigma_p(k)$, qed.

Subtracting this equality from the preceding equality, we obtain

$$\begin{aligned}
&z\left(\sigma_{p-1}\right)-z\left(\sigma_p\right)\\
&=\left(x_p y_p + x_{i_p} y_{m_p} + \sum_{\substack{k\in[n];\\ k\neq p \text{ and } k\neq i_p}} x_k y_{\sigma_p(k)}\right) - \left(x_p y_{m_p} + x_{i_p} y_p + \sum_{\substack{k\in[n];\\ k\neq p \text{ and } k\neq i_p}} x_k y_{\sigma_p(k)}\right)\\
&= x_p y_p + x_{i_p} y_{m_p} - x_p y_{m_p} - x_{i_p} y_p = \left(x_{i_p} - x_p\right)\left(y_{m_p} - y_p\right).
\end{aligned}$$

This proves (942).]

Recall that $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ are numbers, i.e., elements of $\mathbb{A}$, where $\mathbb{A}$ is either $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$. Consider this $\mathbb{A}$. We have $1-1=0\leq n$. Hence, (16) (applied to $u=1$, $v=n$ and $a_s = z(\sigma_s)$) yields

$$\sum_{s=1}^{n}\left(z\left(\sigma_s\right)-z\left(\sigma_{s-1}\right)\right) = z\left(\underbrace{\sigma_n}_{\substack{=\sigma\\ \text{(by (936))}}}\right) - z\left(\underbrace{\sigma_{1-1}}_{\substack{=\sigma_0=\text{id}\\ \text{(by (935))}}}\right) = z\left(\sigma\right) - z\left(\text{id}\right).$$

But

$$\begin{aligned}
&\sum_{p=1}^{n}\left(z\left(\sigma_{p-1}\right)-z\left(\sigma_p\right)\right)\\
&=\sum_{s=1}^{n}\underbrace{\left(z\left(\sigma_{s-1}\right)-z\left(\sigma_s\right)\right)}_{=-\left(z\left(\sigma_s\right)-z\left(\sigma_{s-1}\right)\right)} \qquad \left(\begin{array}{c}\text{here, we have renamed the}\\ \text{summation index } p \text{ as } s \text{ in the sum}\end{array}\right)\\
&=\sum_{s=1}^{n}\left(-\left(z\left(\sigma_s\right)-z\left(\sigma_{s-1}\right)\right)\right) = -\underbrace{\sum_{s=1}^{n}\left(z\left(\sigma_s\right)-z\left(\sigma_{s-1}\right)\right)}_{=z(\sigma)-z(\text{id})}\\
&=-\left(z\left(\sigma\right)-z\left(\text{id}\right)\right) = \underbrace{z\left(\text{id}\right)}_{\substack{=\sum_{k=1}^{n} x_k y_{\text{id}(k)}\\ \text{(by the definition of } z(\text{id}))}} - \underbrace{z\left(\sigma\right)}_{\substack{=\sum_{k=1}^{n} x_k y_{\sigma(k)}\\ \text{(by the definition of } z(\sigma))}}\\
&=\sum_{k=1}^{n} x_k \underbrace{y_{\text{id}(k)}}_{\substack{=y_k\\ \text{(since } \text{id}(k)=k)}} - \sum_{k=1}^{n} x_k y_{\sigma(k)} = \sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k y_{\sigma(k)}.
\end{aligned}$$

Hence,

$$\sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k y_{\sigma(k)}$$
$$= \sum_{p=1}^{n} \underbrace{\left(z\left(\sigma_{p-1}\right) - z\left(\sigma_p\right)\right)}_{\substack{=\left(x_{i_p}-x_p\right)\left(y_{m_p}-y_p\right) \\ \text{(by (942))}}} = \sum_{p=1}^{n} \left(x_{i_p} - x_p\right)\left(y_{m_p} - y_p\right) = \sum_{k=1}^{n} \left(x_{i_k} - x_k\right)\left(y_{m_k} - y_k\right)$$

(here, we have renamed the summation index p as k). This solves Exercise 5.25 **(f)**.

(g) Fix $k \in [n]$. Then, $i_k \in [k]$ (by (934)), so that $i_k \leq k$ and therefore $x_{i_k} \geq x_k$ (since $x_1 \geq x_2 \geq \cdots \geq x_n$). Hence, $x_{i_k} - x_k \geq 0$.

Also, $m_k \in [k]$ (by Exercise 5.25 **(b)**), so that $m_k \leq k$ and thus $y_{m_k} \geq y_k$ (since $y_1 \geq y_2 \geq \cdots \geq y_n$). Hence, $y_{m_k} - y_k \geq 0$. Now,

$$\underbrace{\left(x_{i_k} - x_k\right)}_{\geq 0} \underbrace{\left(y_{m_k} - y_k\right)}_{\geq 0} \geq 0. \tag{946}$$

Now, forget that we fixed k. We thus have proven (946) for each $k \in [n]$. Now, Exercise 5.25 **(f)** yields

$$\sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k y_{\sigma(k)} = \sum_{k=1}^{n} \underbrace{\left(x_{i_k} - x_k\right)\left(y_{m_k} - y_k\right)}_{\substack{\geq 0 \\ \text{(by (946))}}} \geq 0.$$

In other words,

$$\sum_{k=1}^{n} x_k y_k \geq \sum_{k=1}^{n} x_k y_{\sigma(k)}.$$

This solves Exercise 5.25 **(g)**. □

7.65. Solution to Exercise 5.27

Before we come to the solution of Exercise 5.27, we need to lay a lot of groundwork. First, we introduce a notation, which generalizes the definition of $\sigma \times \tau$ in Exercise 5.27:

Definition 7.134. If X, X', Y and Y' are four sets and if $\alpha : X \to X'$ and $\beta : Y \to Y'$ are two maps, then $\alpha \times \beta$ will denote the map

$$X \times Y \to X' \times Y',$$
$$(x, y) \mapsto (\alpha(x), \beta(y)).$$

We shall use Definition 7.134 throughout Section 7.65.

The following properties of this definition are straightforward to prove:

Proposition 7.135. Let X, X', X'', Y, Y' and Y'' be six sets. Let $\alpha : X \to X'$, $\alpha' : X' \to X''$, $\beta : Y \to Y'$ and $\beta' : Y' \to Y''$ be four maps. Then,

$$(\alpha' \times \beta') \circ (\alpha \times \beta) = (\alpha' \circ \alpha) \times (\beta' \circ \beta).$$

Proof of Proposition 7.135. Straightforward computation reveals that the maps $(\alpha' \times \beta') \circ (\alpha \times \beta)$ and $(\alpha' \circ \alpha) \times (\beta' \circ \beta)$ send any given element $(x, y) \in X \times Y$ to the same image (namely, to $(\alpha'(\alpha(x)), \beta'(\beta(y)))$). Thus, these two maps are identical. This proves Proposition 7.135. □

Proposition 7.136. Let U and V be two sets. Then, $\mathrm{id}_U \times \mathrm{id}_V = \mathrm{id}_{U \times V}$.

Proof of Proposition 7.136. Again, this follows from the straightforward computation of $(\mathrm{id}_U \times \mathrm{id}_V)(x, y)$ for each $(x, y) \in U \times V$. □

Corollary 7.137. Let U and V be two sets. Let $k \in \mathbb{N}$. Let $f_1, f_2, \ldots, f_k$ be k maps from U to U. For each $i \in \{1, 2, \ldots, k\}$, define a map $g_i : U \times V \to U \times V$ by $g_i = f_i \times \mathrm{id}_V$. Then,

$$g_1 \circ g_2 \circ \cdots \circ g_k = (f_1 \circ f_2 \circ \cdots \circ f_k) \times \mathrm{id}_V.$$

Proof of Corollary 7.137. We claim that

$$g_1 \circ g_2 \circ \cdots \circ g_m = (f_1 \circ f_2 \circ \cdots \circ f_m) \times \mathrm{id}_V \qquad (947)$$

for each $m \in \{0, 1, \ldots, k\}$.

[*Proof of (947):* We shall prove (947) by induction on m:

Induction base: We have

$$g_1 \circ g_2 \circ \cdots \circ g_0 = (\text{empty composition of maps } U \times V \to U \times V) = \mathrm{id}_{U \times V}.$$

Comparing this with

$$\underbrace{(f_1 \circ f_2 \circ \cdots \circ f_0)}_{\substack{=(\text{empty composition of maps } U \to U) \\ = \mathrm{id}_U}} \times \mathrm{id}_V = \mathrm{id}_U \times \mathrm{id}_V = \mathrm{id}_{U \times V} \qquad \text{(by Proposition 7.136)},$$

we obtain $g_1 \circ g_2 \circ \cdots \circ g_0 = (f_1 \circ f_2 \circ \cdots \circ f_0) \times \mathrm{id}_V$. In other words, (947) holds for $m = 0$. This completes the induction base.

Induction step: Let $M \in \{0, 1, \ldots, k\}$ be positive. Assume that (947) holds for $m = M - 1$. We must prove that (947) holds for $m = M$.

We have assumed that (947) holds for $m = M - 1$. In other words, we have

$$g_1 \circ g_2 \circ \cdots \circ g_{M-1} = (f_1 \circ f_2 \circ \cdots \circ f_{M-1}) \times \mathrm{id}_V.$$

But M is positive; thus,

$$\begin{aligned}
& g_1 \circ g_2 \circ \cdots \circ g_M \\
&= \underbrace{(g_1 \circ g_2 \circ \cdots \circ g_{M-1})}_{=(f_1\circ f_2\circ\cdots\circ f_{M-1})\times \mathrm{id}_V} \circ \underbrace{g_M}_{\substack{=f_M\times \mathrm{id}_V \\ \text{(by the definition of } g_M)}} \\
&= ((f_1 \circ f_2 \circ \cdots \circ f_{M-1}) \times \mathrm{id}_V) \circ (f_M \times \mathrm{id}_V) \\
&= \underbrace{((f_1 \circ f_2 \circ \cdots \circ f_{M-1}) \circ f_M)}_{=f_1\circ f_2\circ\cdots\circ f_M} \times \underbrace{(\mathrm{id}_V \circ \mathrm{id}_V)}_{=\mathrm{id}_V} \\
&\qquad \left(\begin{array}{c} \text{by Proposition 7.135} \\ \text{(applied to } X = U,\ X' = U,\ X'' = U,\ Y = V,\ Y' = V,\ Y'' = V, \\ \alpha = f_M,\ \alpha' = f_1 \circ f_2 \circ \cdots \circ f_{M-1},\ \beta = \mathrm{id}_V \text{ and } \beta' = \mathrm{id}_V) \end{array}\right) \\
&= (f_1 \circ f_2 \circ \cdots \circ f_M) \times \mathrm{id}_V .
\end{aligned}$$

In other words, (947) holds for $m = M$. This completes the induction step. Thus, (947) is proven.]

Now, $k \in \{0, 1, \ldots, k\}$ (since $k \in \mathbb{N}$). Hence, (947) (applied to $m = k$) shows that $g_1 \circ g_2 \circ \cdots \circ g_k = (f_1 \circ f_2 \circ \cdots \circ f_k) \times \mathrm{id}_V$. This proves Corollary 7.137. □

Corollary 7.138. Let X, X', Y and Y' be four sets. Let $\alpha : X \to X'$ and $\beta : Y \to Y'$ be two bijective maps. Then, the map $\alpha \times \beta : X \times Y \to X' \times Y'$ is bijective as well, and its inverse is the map $\alpha^{-1} \times \beta^{-1} : X' \times Y' \to X \times Y$.

Proof of Corollary 7.138. This can be derived from Proposition 7.136 and Proposition 7.135 (or, again, checked by straightforward computation). □

Corollary 7.139. Let U and V be two sets. Let σ be a permutation of U. Let τ be a permutation of V. Then, $\sigma \times \tau$ is a permutation of $U \times V$.

Proof of Corollary 7.139. The map σ is a permutation of U. In other words, σ is a bijective map $U \to U$. Similarly, τ is a bijective map $V \to V$. Hence, Corollary 7.138 (applied to $X = U$, $Y = U$, $Y = V$, $Y' = V$, $\alpha = \sigma$ and $\beta = \tau$) shows that the map $\sigma \times \tau : U \times V \to U \times V$ is bijective as well. In other words, $\sigma \times \tau$ is a permutation of $U \times V$. This proves Corollary 7.139. □

Proposition 7.140. Let U and V be two sets. Let $\gamma : U \times V \to V \times U$ be the map defined by

$$(\gamma((u, v)) = (v, u) \qquad \text{for each } (u, v) \in U \times V).$$

Then:

(a) The map $\gamma : U \times V \to V \times U$ is a bijection.

(b) Let f be any map $U \to U$. Let g be any map $V \to V$. Then, $f \times g = \gamma^{-1} \circ (g \times f) \circ \gamma$.

Proof of Proposition 7.140. **(a)** Let $\delta : V \times U \to U \times V$ be the map defined by

$$(\delta((v,u)) = (u,v) \qquad \text{for each } (v,u) \in V \times U).$$

Clearly, the maps γ and δ are mutually inverse. Thus, the map γ is invertible, i.e., is a bijection. This proves Proposition 7.140 **(a)**.

(b) For each $(u,v) \in U \times V$, we have

$$\begin{aligned}
&(\gamma \circ (f \times g))((u,v)) \\
&= \gamma \left(\underbrace{(f \times g)((u,v))}_{\substack{=(f(u),g(v)) \\ \text{(by the definition of } f \times g)}} \right) \\
&= \gamma((f(u), g(v))) = (g(v), f(u)) \qquad \text{(by the definition of } \gamma)
\end{aligned}$$

and

$$\begin{aligned}
&((g \times f) \circ \gamma)((u,v)) \\
&= (g \times f) \left(\underbrace{\gamma((u,v))}_{=(v,u)} \right) \\
&= (g \times f)((v,u)) = (g(v), f(u)) \qquad \text{(by the definition of } g \times f).
\end{aligned}$$

Comparing these two equalities, we conclude that

$$(\gamma \circ (f \times g))((u,v)) = ((g \times f) \circ \gamma)((u,v)) \qquad \text{for each } (u,v) \in U \times V.$$

In other words, $\gamma \circ (f \times g) = (g \times f) \circ \gamma$. Thus,

$$\gamma^{-1} \circ \underbrace{(g \times f) \circ \gamma}_{=\gamma \circ (f \times g)} = \underbrace{\gamma^{-1} \circ \gamma}_{=\mathrm{id}_{U \times V}} \circ (f \times g) = f \times g.$$

This proves Proposition 7.140 **(b)**. ☐

Next, let us discuss some properties of signs of permutations of finite sets:

Proposition 7.141. Let X and Y be two finite sets. Let $f : X \to Y$ be a bijection. Let σ be a permutation of X. Then, $f \circ \sigma \circ f^{-1}$ is a permutation of Y and satisfies $(-1)^{f \circ \sigma \circ f^{-1}} = (-1)^{\sigma}$.

Proof of Proposition 7.141. The map σ is a permutation of X. In other words, σ is a bijection from X to X. In other words, σ is a bijective map $X \to X$.

Also, $f : X \to Y$ is a bijection. Hence, the inverse f^{-1} of f is well-defined and is a bijection $Y \to X$.

The map $f \circ \sigma \circ f^{-1} : Y \to Y$ is a bijection (since it is the composition of the bijections f^{-1}, σ and f). In other words, $f \circ \sigma \circ f^{-1}$ is a permutation of Y. Hence, $(-1)^{f\circ\sigma\circ f^{-1}}$ is well-defined.

It remains to prove that $(-1)^{f\circ\sigma\circ f^{-1}} = (-1)^{\sigma}$.

Define $n \in \mathbb{N}$ by $n = |X|$. (This is well-defined, since the set X is finite.)

We have $|X| = n = |\{1, 2, \ldots, n\}|$. Hence, there exists a bijection $\psi : X \to \{1, 2, \ldots, n\}$. Consider such a ψ.

Recall the definition of $(-1)^{\sigma}_{\psi}$ given in Exercise 5.12. This definition shows that $(-1)^{\sigma}_{\psi} = (-1)^{\psi\circ\sigma\circ\psi^{-1}}$.

The map $\psi \circ f^{-1}$ is a bijection from Y to $\{1, 2, \ldots, n\}$ (since ψ and f^{-1} are bijections). Hence, the definition of $(-1)^{f\circ\sigma\circ f^{-1}}_{\psi\circ f^{-1}}$ (given in Exercise 5.12) shows that

$$(-1)^{f\circ\sigma\circ f^{-1}}_{\psi\circ f^{-1}} = (-1)^{(\psi\circ f^{-1})\circ(f\circ\sigma\circ f^{-1})\circ(\psi\circ f^{-1})^{-1}}.$$

In view of

$$\underbrace{\left(\psi \circ f^{-1}\right) \circ \left(f \circ \sigma \circ f^{-1}\right)}_{=\psi\circ f^{-1}\circ f\circ\sigma\circ f^{-1}} \circ \underbrace{\left(\psi \circ f^{-1}\right)^{-1}}_{=\left(f^{-1}\right)^{-1}\circ\psi^{-1}}$$
$$= \psi \circ \underbrace{f^{-1} \circ f}_{=\mathrm{id}_X} \circ \sigma \circ \underbrace{f^{-1} \circ \left(f^{-1}\right)^{-1}}_{=\mathrm{id}_X} \circ \psi^{-1} = \psi \circ \sigma \circ \psi^{-1},$$

this rewrites as $(-1)^{f\circ\sigma\circ f^{-1}}_{\psi\circ f^{-1}} = (-1)^{\psi\circ\sigma\circ\psi^{-1}}$.

But the definition of $(-1)^{\sigma}$ (given in Exercise 5.12) shows that $(-1)^{\sigma} = (-1)^{\sigma}_{\phi}$ for any bijection $\phi : X \to \{1, 2, \ldots, n\}$. Applying this to $\phi = \psi$, we obtain

$$(-1)^{\sigma} = (-1)^{\sigma}_{\psi} = (-1)^{\psi\circ\sigma\circ\psi^{-1}}. \tag{948}$$

Also, the definition of $(-1)^{f\circ\sigma\circ f^{-1}}$ (given in Exercise 5.12) shows that $(-1)^{f\circ\sigma\circ f^{-1}} = (-1)^{f\circ\sigma\circ f^{-1}}_{\phi}$ for any bijection $\phi : Y \to \{1, 2, \ldots, n\}$. Applying this to $\phi = \psi \circ f^{-1}$, we obtain

$$(-1)^{f\circ\sigma\circ f^{-1}} = (-1)^{f\circ\sigma\circ f^{-1}}_{\psi\circ f^{-1}} = (-1)^{\psi\circ\sigma\circ\psi^{-1}}.$$

Comparing this with (948), we obtain $(-1)^{f\circ\sigma\circ f^{-1}} = (-1)^{\sigma}$. This completes the proof of Proposition 7.141. □

Corollary 7.142. Let U and V be two finite sets. Let σ be a permutation of U. Let τ be a permutation of V. Then, $(-1)^{\sigma\times\tau} = (-1)^{\tau\times\sigma}$.

Proof of Corollary 7.142. Corollary 7.139 yields that $\sigma \times \tau$ is a permutation of $U \times V$. Hence, $(-1)^{\sigma \times \tau}$ is well-defined.

Consider the map $\gamma : U \times V \to V \times U$ defined in Proposition 7.140. Then, Proposition 7.140 **(a)** shows that this map γ is a bijection. Hence, Proposition 7.141 (applied to $U \times V$, $V \times U$, γ and $\sigma \times \tau$ instead of X, Y, f and σ) shows that $\gamma \circ (\sigma \times \tau) \circ \gamma^{-1}$ is a permutation of $V \times U$ and satisfies $(-1)^{\gamma \circ (\sigma \times \tau) \circ \gamma^{-1}} = (-1)^{\sigma \times \tau}$.

But Proposition 7.140 **(b)** (applied to $f = \sigma$ and $g = \tau$) shows that $\sigma \times \tau = \gamma^{-1} \circ (\tau \times \sigma) \circ \gamma$. Hence,

$$\gamma \circ \underbrace{(\sigma \times \tau)}_{= \gamma^{-1} \circ (\tau \times \sigma) \circ \gamma} \circ \gamma^{-1} = \underbrace{\gamma \circ \gamma^{-1}}_{= \mathrm{id}_{V \times U}} \circ (\tau \times \sigma) \circ \underbrace{\gamma \circ \gamma^{-1}}_{= \mathrm{id}_{V \times U}} = \tau \times \sigma.$$

Thus, $(-1)^{\gamma \circ (\sigma \times \tau) \circ \gamma^{-1}} = (-1)^{\tau \times \sigma}$. Comparing this with $(-1)^{\gamma \circ (\sigma \times \tau) \circ \gamma^{-1}} = (-1)^{\sigma \times \tau}$, we obtain $(-1)^{\sigma \times \tau} = (-1)^{\tau \times \sigma}$. This proves Corollary 7.142. □

For the rest of Section 7.65, we shall use Definition 5.36.

Proposition 7.143. Let X be a set. Let $m \in \mathbb{N}$. Let M be the set $\{1, 2, \ldots, m\}$. Let x and y be two distinct elements of X. Then, $t_{(x,1),(y,1)}, t_{(x,2),(y,2)}, \ldots, t_{(x,m),(y,m)}$ are m well-defined transpositions of the set $X \times M$, and we have

$$t_{x,y} \times \mathrm{id}_M = t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}.$$

Proof of Proposition 7.143. Let us first give a quick informal proof of Proposition 7.143:

For each $j \in \{1, 2, \ldots, m\}$, the transposition $t_{(x,j),(y,j)}$ switches (x, j) with (y, j) while leaving all other elements of $X \times M$ unchanged. Thus, the composition $t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}$ of these m transpositions $t_{(x,m),(y,m)}, t_{(x,m-1),(y,m-1)}, \ldots, t_{(x,1),(y,1)}$ switches the elements $(x, 1), (x, 2), \ldots, (x, m)$ with the elements $(y, 1), (y, 2), \ldots, (y, m)$, respectively, while leaving all other elements of $X \times M$ unchanged[442]. In other words, $t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}$ switches each element of the form (x, q) (with $q \in \{1, 2, \ldots, m\}$) with the corresponding (y, q), while leaving all other elements of $X \times M$ unchanged. In

[442]Here we are using the observation that the $2m$ elements $(x, 1), (y, 1), (x, 2), (y, 2), \ldots, (x, m), (y, m)$ of $X \times M$ are distinct (since x and y are distinct), whence any element of $X \times M$ is moved (i.e., not left unchanged) by **at most one** of the m transpositions $t_{(x,m),(y,m)}, t_{(x,m-1),(y,m-1)}, \ldots, t_{(x,1),(y,1)}$.

other words,

$$\left(t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}\right)((u,q)) = \begin{cases} (y,q), & \text{if } u = x; \\ (x,q), & \text{if } u = y; \\ (u,q), & \text{otherwise} \end{cases} \tag{949}$$

for each $(u,q) \in X \times M$.

But the transposition $t_{x,y}$ switches x with y while leaving all other elements of X unchanged. Thus,

$$t_{x,y}(u) = \begin{cases} y, & \text{if } u = x; \\ x, & \text{if } u = y; \\ u, & \text{otherwise} \end{cases} \tag{950}$$

for each $u \in X$. Hence, for each $(u,q) \in X \times M$, we have

$$\begin{aligned} \left(t_{x,y} \times \operatorname{id}_M\right)((u,q)) &= \left(t_{x,y}(u), \underbrace{\operatorname{id}_M(q)}_{=q}\right) = \left(t_{x,y}(u), q\right) \\ &= \left(\begin{cases} y, & \text{if } u = x; \\ x, & \text{if } u = y; \\ u, & \text{otherwise} \end{cases}, q\right) \qquad \text{(by (950))} \\ &= \begin{cases} (y,q), & \text{if } u = x; \\ (x,q), & \text{if } u = y; \\ (u,q), & \text{otherwise} \end{cases} \\ &= \left(t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}\right)((u,q)) \end{aligned}$$

(by (949)). In other words, we have

$$t_{x,y} \times \operatorname{id}_M = t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}.$$

Thus, Proposition 7.143 follows.

Let us now show how to prove Proposition 7.143 rigorously. The following rigorous proof is, of course, just a formalization of the above informal argument.

We have $x \neq y$ (since x and y are distinct). Thus, for each $p \in \{1,2,\ldots,m\}$, we have $(x,p) \neq (y,p)$. Hence, for each $p \in \{1,2,\ldots,m\}$, the transposition $t_{(x,p),(y,p)}$ of $X \times M$ is well-defined. In other words, $t_{(x,1),(y,1)}, t_{(x,2),(y,2)}, \ldots, t_{(x,m),(y,m)}$ are m well-defined transpositions of the set $X \times M$.

It remains to prove that $t_{x,y} \times \operatorname{id}_M = t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}$.

Let us first observe the following fact: For any $j \in \{1,2,\ldots,m\}$ and $q \in M$ and $z \in X$ satisfying $j \neq q$, we have

$$t_{(x,j),(y,j)}((z,q)) = (z,q). \tag{951}$$

[*Proof of (951):* Let $j \in \{1, 2, \ldots, m\}$ and $q \in M$ and $z \in X$ be such that $j \neq q$.

The pair (z, q) is distinct from both (x, j) and (y, j) (since $j \neq q$). In other words, $(z, q) \notin \{(x, j), (y, j)\}$. Combining this with $(z, q) \in X \times M$ (since $z \in X$ and $q \in M$), we obtain $(z, q) \in (X \times M) \setminus \{(x, j), (y, j)\}$. Hence, Lemma 7.93 **(c)** (applied to $X \times M$, (x, j), (y, j) and (z, q) instead of X, i, j and k) yields $t_{(x,j),(y,j)}((z, q)) = (z, q)$. This proves (951).]

Define a map $s : X \times M \to X \times M$ by

$$s = t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}. \tag{952}$$

Fix $c \in X \times M$. We shall show that $(t_{x,y} \times \mathrm{id}_M)(c) = s(c)$.

Write the element $c \in X \times M$ in the form $c = (z, q)$ for some $z \in X$ and $q \in M$.

Applying the map $t_{x,y} \times \mathrm{id}_M$ to both sides of the equality $c = (z, q)$, we obtain

$$\begin{aligned}(t_{x,y} \times \mathrm{id}_M)(c) &= (t_{x,y} \times \mathrm{id}_M)((z, q)) = \left(t_{x,y}(z), \underbrace{\mathrm{id}_M(q)}_{=q}\right) \\ &= (t_{x,y}(z), q). \end{aligned} \tag{953}$$

We have $q \in M = \{1, 2, \ldots, m\}$. We are in one of the following three cases:

Case 1: We have $z = x$.

Case 2: We have $z = y$.

Case 3: We have neither $z = x$ nor $z = y$.

Let us first consider Case 1. In this case, we have $z = x$. But Lemma 7.93 **(a)** (applied to x and y instead of i and j) yields $t_{x,y}(x) = y$. But $t_{x,y}\left(\underbrace{z}_{=x}\right) = t_{x,y}(x) = y$.

Hence, (953) yields

$$(t_{x,y} \times \mathrm{id}_M)(c) = \left(\underbrace{t_{x,y}(z)}_{=y}, q\right) = (y, q). \tag{954}$$

On the other hand, $(x, q) \neq (y, q)$ (since $x \neq y$). Hence, the two elements (x, q) and (y, q) of $X \times M$ are distinct. Thus, Lemma 7.93 **(a)** (applied to $X \times M$, (x, q) and (y, q) instead of X, i and j) yields

$$t_{(x,q),(y,q)}((x, q)) = (y, q). \tag{955}$$

Next, we observe that

$$t_{(x,j),(y,j)}((x, q)) = (x, q) \qquad \text{for each } j \in \{1, 2, \ldots, m\} \text{ satisfying } j < q. \tag{956}$$

[*Proof of (956):* Let $j \in \{1, 2, \ldots, m\}$ be such that $j < q$. Thus, $j \neq q$. Hence, (951) (applied to $z = x$) yields $t_{(x,j),(y,j)}((x, q)) = (x, q)$. This proves (956).]

Next, we observe that

$$t_{(x,j),(y,j)}((y,q)) = (y,q) \qquad \text{for each } j \in \{1,2,\ldots,m\} \text{ satisfying } j > q. \tag{957}$$

[*Proof of (957):* Let $j \in \{1,2,\ldots,m\}$ be such that $j > q$. Thus, $j \neq q$. Hence, (951) (applied to $z = y$) yields $t_{(x,j),(y,j)}((y,q)) = (y,q)$. This proves (957).]

Now, we can apply Lemma 7.102 to $X \times M$, $t_{(x,j),(y,j)}$, (x,q), (y,q) and q instead of X, f_j, x, y and i (since the equalities (955), (956) and (957) hold). Thus, we obtain

$$\left(t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}\right)((x,q)) = (y,q). \tag{958}$$

In view of (952), this rewrites as $s((x,q)) = (y,q)$. Comparing this with (954), we obtain $\left(t_{x,y} \times \mathrm{id}_M\right)(c) = s((x,q))$. Comparing this with $s\left(\underbrace{c}_{=(z,q)}\right) = s\left(\left(\underbrace{z}_{=x}, q\right)\right) = s((x,q))$, we obtain $\left(t_{x,y} \times \mathrm{id}_M\right)(c) = s(c)$. Hence, $\left(t_{x,y} \times \mathrm{id}_M\right)(c) = s(c)$ is proven in Case 1.

We leave it to the reader to verify $\left(t_{x,y} \times \mathrm{id}_M\right)(c) = s(c)$ in Case 2. (The verification is analogous to what we did in Case 1, but the roles of x and y are often interchanged, and instead of Lemma 7.93 **(a)** we must now apply Lemma 7.93 **(b)**.)

Let us finally consider Case 3. In this case, we have neither $z = x$ nor $z = y$. Thus, $z \notin \{x,y\}$. Combining this with $z \in X$, we obtain $z \in X \setminus \{x,y\}$. Thus, Lemma 7.93 **(c)** (applied to x, y and z instead of i, j and k) yields $t_{x,y}(z) = z$.

Hence, (953) yields

$$\left(t_{x,y} \times \mathrm{id}_M\right)(c) = \left(\underbrace{t_{x,y}(z)}_{=z}, q\right) = (z,q). \tag{959}$$

Next, we observe that

$$t_{(x,j),(y,j)}((z,q)) = (z,q) \qquad \text{for each } j \in \{1,2,\ldots,m\}. \tag{960}$$

[*Proof of (960):* Let $j \in \{1,2,\ldots,m\}$. Thus, $j \in \{1,2,\ldots,m\} = M$. The elements (x,j) and (y,j) of $X \times M$ are distinct (since $x \neq y$). The element (z,q) is distinct from both (x,j) and (y,j) (since we have neither $z = x$ nor $z = y$). In other words, $(z,q) \notin \{(x,j),(y,j)\}$. Combining this with $(z,q) \in X \times M$, we obtain $(z,q) \in (X \times M) \setminus \{(x,j),(y,j)\}$. Hence, Lemma 7.93 **(c)** (applied to $X \times M$, (x,j), (y,j) and (z,q) instead of X, i, j and k) yields $t_{(x,j),(y,j)}((z,q)) = (z,q)$. This proves (960).]

Now, we can apply Lemma 7.101 to $X \times M$, $t_{(x,j),(y,j)}$ and (z,q) instead of X, f_j and x (since (960) holds). Thus, we obtain

$$\left(t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}\right)((z,q)) = (z,q). \tag{961}$$

In view of (952), this rewrites as $s((z,q)) = (z,q)$. Comparing this with (959), we obtain $(t_{x,y} \times \mathrm{id}_M)(c) = s((z,q))$. Comparing this with $s\left(\underbrace{c}_{=(z,q)}\right) = s((z,q))$, we obtain $(t_{x,y} \times \mathrm{id}_M)(c) = s(c)$. Hence, $(t_{x,y} \times \mathrm{id}_M)(c) = s(c)$ is proven in Case 3.

We have now proven $(t_{x,y} \times \mathrm{id}_M)(c) = s(c)$ in each of the three Cases 1, 2 and 3. Thus, $(t_{x,y} \times \mathrm{id}_M)(c) = s(c)$ always holds.

Now, forget that we fixed c. We thus have proven that $(t_{x,y} \times \mathrm{id}_M)(c) = s(c)$ for each $c \in X \times M$. Thus,

$$t_{x,y} \times \mathrm{id}_M = s = t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}.$$

This completes the rigorous proof of Proposition 7.143. □

Corollary 7.144. Let X be a finite set. Let $m \in \mathbb{N}$. Let M be the set $\{1, 2, \ldots, m\}$. Let x and y be two distinct elements of X. Then, the permutation $t_{x,y} \times \mathrm{id}_M$ of $X \times M$ satisfies $(-1)^{t_{x,y} \times \mathrm{id}_M} = (-1)^m$.

Proof of Corollary 7.144. Corollary 7.139 (applied to X, M, $t_{x,y}$ and id_M instead of U, V, σ and τ) shows that $t_{x,y} \times \mathrm{id}_M$ is a permutation of $X \times M$. Proposition 7.143 shows that $t_{(x,1),(y,1)}, t_{(x,2),(y,2)}, \ldots, t_{(x,m),(y,m)}$ are m well-defined transpositions of the set $X \times M$, and that we have

$$t_{x,y} \times \mathrm{id}_M = t_{(x,m),(y,m)} \circ t_{(x,m-1),(y,m-1)} \circ \cdots \circ t_{(x,1),(y,1)}. \quad (962)$$

The equality (962) shows that $t_{x,y} \times \mathrm{id}_M$ can be written as a composition of m transpositions of $X \times M$ (since $t_{(x,1),(y,1)}, t_{(x,2),(y,2)}, \ldots, t_{(x,m),(y,m)}$ are m transpositions of $X \times M$). Hence, Corollary 7.100 (applied to $X \times M$, $t_{x,y} \times \mathrm{id}_M$ and m instead of X, σ and k) shows that $(-1)^{t_{x,y} \times \mathrm{id}_M} = (-1)^m$. This proves Corollary 7.144. □

Corollary 7.145. Let U and V be two finite sets. Let i and j be two distinct elements of U. Then, the permutation $t_{i,j} \times \mathrm{id}_V$ of $U \times V$ satisfies $(-1)^{t_{i,j} \times \mathrm{id}_V} = (-1)^{|V|}$.

Proof of Corollary 7.145. Define an $m \in \mathbb{N}$ by $m = |V|$. (This is well-defined, since the set V is finite.)

Let M be the set $\{1, 2, \ldots, m\}$. Thus, $M = \{1, 2, \ldots, m\}$, so that $|M| = |\{1, 2, \ldots, m\}| = m = |V|$. The sets V and M are finite and have the same size (since $|M| = |V|$). Hence, there exists a bijection $\phi : M \to V$. Consider such a ϕ.

Corollary 7.138 (applied to U, U, M, V, id_U and ϕ instead of X, X', Y, Y', α and β) shows that the map $\mathrm{id}_U \times \phi : U \times M \to U \times V$ is bijective. In other words, $\mathrm{id}_U \times \phi$ is a bijection. Also, Corollary 7.139 (applied to M, $t_{i,j}$ and id_M instead of V, σ and τ) shows that $t_{i,j} \times \mathrm{id}_M$ is a permutation of $U \times M$. Similarly, $t_{i,j} \times \mathrm{id}_V$ is a permutation of $U \times V$.

Now, Proposition 7.141 (applied to $U \times M$, $U \times V$, $\operatorname{id}_U \times \phi$ and $t_{i,j} \times \operatorname{id}_M$ instead of X, Y, f and σ) shows that $(\operatorname{id}_U \times \phi) \circ (t_{i,j} \times \operatorname{id}_M) \circ (\operatorname{id}_U \times \phi)^{-1}$ is a permutation of $U \times V$ and satisfies

$$(-1)^{(\operatorname{id}_U \times \phi) \circ (t_{i,j} \times \operatorname{id}_M) \circ (\operatorname{id}_U \times \phi)^{-1}} = (-1)^{t_{i,j} \times \operatorname{id}_M}. \tag{963}$$

But Corollary 7.144 (applied to $X = U$, $x = i$ and $y = j$) shows that the permutation $t_{i,j} \times \operatorname{id}_M$ of $U \times M$ satisfies

$$(-1)^{t_{i,j} \times \operatorname{id}_M} = (-1)^m. \tag{964}$$

But Proposition 7.135 (applied to U, U, U, M, M, V, $t_{i,j}$, id_U, id_M and ϕ instead of X, X', X'', Y, Y', Y'', α, α', β and β') yields

$$(\operatorname{id}_U \times \phi) \circ (t_{i,j} \times \operatorname{id}_M) = \underbrace{(\operatorname{id}_U \circ t_{i,j})}_{= t_{i,j}} \times \underbrace{(\phi \circ \operatorname{id}_M)}_{= \phi} = t_{i,j} \times \phi. \tag{965}$$

On the other hand, Proposition 7.135 (applied to U, U, U, M, V, V, id_U, $t_{i,j}$, ϕ and id_V instead of X, X', X'', Y, Y', Y'', α, α', β and β') yields

$$(t_{i,j} \times \operatorname{id}_V) \circ (\operatorname{id}_U \times \phi) = \underbrace{(t_{i,j} \circ \operatorname{id}_U)}_{= t_{i,j}} \times \underbrace{(\operatorname{id}_V \circ \phi)}_{= \phi} = t_{i,j} \times \phi.$$

Comparing this with (965), we obtain

$$(\operatorname{id}_U \times \phi) \circ (t_{i,j} \times \operatorname{id}_M) = (t_{i,j} \times \operatorname{id}_V) \circ (\operatorname{id}_U \times \phi).$$

Hence,

$$\begin{aligned} &\underbrace{(\operatorname{id}_U \times \phi) \circ (t_{i,j} \times \operatorname{id}_M)}_{= (t_{i,j} \times \operatorname{id}_V) \circ (\operatorname{id}_U \times \phi)} \circ (\operatorname{id}_U \times \phi)^{-1} \\ &= (t_{i,j} \times \operatorname{id}_V) \circ \underbrace{(\operatorname{id}_U \times \phi) \circ (\operatorname{id}_U \times \phi)^{-1}}_{= \operatorname{id}_{U \times V}} = t_{i,j} \times \operatorname{id}_V. \end{aligned}$$

Thus, (963) rewrites as $(-1)^{t_{i,j} \times \operatorname{id}_V} = (-1)^{t_{i,j} \times \operatorname{id}_M}$. Thus,

$$\begin{aligned} (-1)^{t_{i,j} \times \operatorname{id}_V} &= (-1)^{t_{i,j} \times \operatorname{id}_M} = (-1)^m \qquad &&(\text{by (964)}) \\ &= (-1)^{|V|} \qquad &&(\text{since } m = |V|). \end{aligned}$$

This proves Corollary 7.145. $\square$

Corollary 7.146. Let U and V be two finite sets. Let σ be a permutation of U. Then, the permutation $\sigma \times \mathrm{id}_V$ of $U \times V$ satisfies $(-1)^{\sigma \times \mathrm{id}_V} = \left((-1)^{\sigma}\right)^{|V|}$.

Proof of Corollary 7.146. Corollary 7.139 (applied to $\tau = \mathrm{id}_V$) shows that $\sigma \times \mathrm{id}_V$ is a permutation of $U \times V$.

Proposition 7.98 (applied to U and σ instead of X and τ) shows that σ can be written as a composition of finitely many transpositions of U. In other words, there exists some $k \in \mathbb{N}$ such that σ can be written as a composition of k transpositions of U. Consider this k. Corollary 7.100 (applied to $X = U$) yields

$$(-1)^{\sigma} = (-1)^{k}. \tag{966}$$

We know that σ can be written as a composition of k transpositions of U. In other words, there exist k transpositions $f_1, f_2, \ldots, f_k$ of U satisfying $\sigma = f_1 \circ f_2 \circ \cdots \circ f_k$. Consider these $f_1, f_2, \ldots, f_k$.

For each $i \in \{1, 2, \ldots, k\}$, define a map $g_i : U \times V \to U \times V$ by $g_i = f_i \times \mathrm{id}_V$.

Let $p \in \{1, 2, \ldots, k\}$. Thus, $g_p = f_p \times \mathrm{id}_V$ (by the definition of g_p). Corollary 7.139 (applied to f_p and id_V instead of σ and τ) shows that $f_p \times \mathrm{id}_V$ is a permutation of $U \times V$. In other words, g_p is a permutation of $U \times V$ (since $g_p = f_p \times \mathrm{id}_V$).

The map f_p is a transposition of U. In other words, $f_p = t_{i,j}$ for some two distinct elements i and j of U (by the definition of a transposition of U). Consider these i and j. Corollary 7.145 shows that the permutation $t_{i,j} \times \mathrm{id}_V$ of $U \times V$ satisfies $(-1)^{t_{i,j} \times \mathrm{id}_V} = (-1)^{|V|}$. Now, $f_p = t_{i,j}$. Thus, $f_p \times \mathrm{id}_V = t_{i,j} \times \mathrm{id}_V$. Hence, $g_p = f_p \times \mathrm{id}_V = t_{i,j} \times \mathrm{id}_V$. Therefore,

$$(-1)^{g_p} = (-1)^{t_{i,j} \times \mathrm{id}_V} = (-1)^{|V|}. \tag{967}$$

Now, forget that we fixed p. We thus have proven that for each $p \in \{1, 2, \ldots, k\}$, the map g_p is a permutation of $U \times V$ and satisfies (967).

Hence, the maps $g_1, g_2, \ldots, g_k$ are k permutations of $U \times V$. Thus, Proposition 7.99 (applied to $X = U \times V$ and $\sigma_h = g_h$) yields

$$(-1)^{g_1 \circ g_2 \circ \cdots \circ g_k} = (-1)^{g_1} \cdot (-1)^{g_2} \cdot \cdots \cdot (-1)^{g_k}. \tag{968}$$

Corollary 7.137 yields

$$g_1 \circ g_2 \circ \cdots \circ g_k = \underbrace{(f_1 \circ f_2 \circ \cdots \circ f_k)}_{\substack{=\sigma \\ (\text{since } \sigma = f_1 \circ f_2 \circ \cdots \circ f_k)}} \times \mathrm{id}_V = \sigma \times \mathrm{id}_V.$$

Thus, $\sigma \times \mathrm{id}_V = g_1 \circ g_2 \circ \cdots \circ g_k$. Therefore,

$$\begin{aligned}(-1)^{\sigma \times \mathrm{id}_V} &= (-1)^{g_1 \circ g_2 \circ \cdots \circ g_k} = (-1)^{g_1} \cdot (-1)^{g_2} \cdot \cdots \cdot (-1)^{g_k} \\ &= \prod_{p=1}^{k} \underbrace{(-1)^{g_p}}_{\substack{=(-1)^{|V|} \\ \text{(by (967))}}} = \prod_{p=1}^{k} (-1)^{|V|} = \left((-1)^{|V|}\right)^k \\ &= (-1)^{|V| \cdot k} = (-1)^{k \cdot |V|} = \left(\underbrace{(-1)^k}_{\substack{=(-1)^{\sigma} \\ \text{(by (966))}}} \right)^{|V|} = \left((-1)^{\sigma}\right)^{|V|}.\end{aligned}$$

This proves Corollary 7.146. □

Corollary 7.147. Let U and V be two finite sets. Let τ be a permutation of V. Then, the permutation $\mathrm{id}_U \times \tau$ of $U \times V$ satisfies $(-1)^{\mathrm{id}_U \times \tau} = \left((-1)^{\tau}\right)^{|U|}$.

Proof of Corollary 7.147. The map id_U is a permutation of U, whereas τ is a permutation of V. Hence, $\mathrm{id}_U \times \tau$ is a permutation of $U \times V$ (by Corollary 7.139 (applied to $\sigma = \mathrm{id}_U$)).

Corollary 7.142 (applied to $\sigma = \mathrm{id}_U$) yields $(-1)^{\mathrm{id}_U \times \tau} = (-1)^{\tau \times \mathrm{id}_U}$. But Corollary 7.146 (applied to V, U and τ instead of U, V and σ) shows that the permutation $\tau \times \mathrm{id}_U$ of $V \times U$ satisfies $(-1)^{\tau \times \mathrm{id}_U} = \left((-1)^{\tau}\right)^{|U|}$. Hence, $(-1)^{\mathrm{id}_U \times \tau} = (-1)^{\tau \times \mathrm{id}_U} = \left((-1)^{\tau}\right)^{|U|}$. This proves Corollary 7.147. □

Corollary 7.148. Let U and V be two finite sets. Let σ be a permutation of U. Let τ be a permutation of V. Then, the permutation $\sigma \times \tau$ of $U \times V$ satisfies $\sigma \times \tau = (\sigma \times \mathrm{id}_V) \circ (\mathrm{id}_U \times \tau)$ and $(-1)^{\sigma \times \tau} = \left((-1)^{\sigma}\right)^{|V|} \left((-1)^{\tau}\right)^{|U|}$.

Proof of Corollary 7.148. Corollary 7.139 (applied to $\sigma = \mathrm{id}_U$) shows that $\mathrm{id}_U \times \tau$ is a permutation of $U \times V$. Similarly, $\sigma \times \mathrm{id}_V$ is a permutation of $U \times V$.

Proposition 7.135 (applied to U, U, U, V, V, V, id_U, σ, τ and id_V instead of X, X', X'', Y, Y', Y'', α, α', β and β') shows that

$$(\sigma \times \mathrm{id}_V) \circ (\mathrm{id}_U \times \tau) = \underbrace{(\sigma \circ \mathrm{id}_U)}_{=\sigma} \times \underbrace{(\mathrm{id}_V \circ \tau)}_{=\tau} = \sigma \times \tau.$$

Thus, $\sigma \times \tau = (\sigma \times \mathrm{id}_V) \circ (\mathrm{id}_U \times \tau)$. Hence,

$$(-1)^{\sigma\times\tau} = (-1)^{(\sigma\times\mathrm{id}_V)\circ(\mathrm{id}_U\times\tau)} = \underbrace{(-1)^{\sigma\times\mathrm{id}_V}}_{\substack{=\left((-1)^{\sigma}\right)^{|V|} \\ \text{(by Corollary 7.146)}}} \cdot \underbrace{(-1)^{\mathrm{id}_U\times\tau}}_{\substack{=\left((-1)^{\tau}\right)^{|U|} \\ \text{(by Corollary 7.147)}}}$$
$$\left(\begin{array}{c}\text{by Exercise 5.12 (c) (applied to } U\times V\text{, } \sigma\times\mathrm{id}_V \\ \text{and } \mathrm{id}_U\times\tau \text{ instead of } X\text{, } \sigma \text{ and } \tau\text{)}\end{array}\right)$$
$$= \left((-1)^{\sigma}\right)^{|V|} \left((-1)^{\tau}\right)^{|U|}.$$

This completes the proof of Corollary 7.148. □

Solution to Exercise 5.27. Exercise 5.27 **(a)** follows from Corollary 7.139. Parts **(b)** and **(c)** of Exercise 5.27 follow from Corollary 7.148. □

7.66. Solution to Exercise 5.28

Throughout Section 7.66, we shall use the same notations that were used in Section 5.8. We shall also use Definition 3.48.

We begin with several minor results that we shall use in our solution of Exercise 5.28. The first of these is a well-known inequality, known as the *triangle inequality for numbers*:

Theorem 7.149. Let x and y be two rational numbers (or real numbers). Then, $|x| + |y| \geq |x+y|$.

Proof of Theorem 7.149. Every rational number (or real number) z satisfies

$$|z| = \max\{z, -z\} \geq z \tag{969}$$

and

$$|z| = \max\{z, -z\} \geq -z. \tag{970}$$

Now, $x+y$ is a rational (or real) number, and thus is either ≥ 0 or < 0. Hence, we are in one of the following two cases:

Case 1: We have $x+y \geq 0$.

Case 2: We have $x+y < 0$.

Let us first consider Case 1. In this case, we have $x+y \geq 0$. Hence, $|x+y| = x+y$. Now, $\underbrace{|x|}_{\substack{\geq x \\ \text{(by (969))}}} + \underbrace{|y|}_{\substack{\geq y \\ \text{(by (969))}}} \geq x+y = |x+y|$. Hence, Theorem 7.149 is proven in Case 1.

Let us now consider Case 2. In this case, we have $x+y < 0$. Hence, $|x+y| = -(x+y) = (-x)+(-y)$. Now, $\underbrace{|x|}_{\substack{\geq -x \\ \text{(by (970))}}} + \underbrace{|y|}_{\substack{\geq -y \\ \text{(by (970))}}} \geq (-x)+(-y) = |x+y|$. Hence, Theorem 7.149 is proven in Case 2.

We have now proven Theorem 7.149 in each of the two Cases 1 and 2. Hence, Theorem 7.149 always holds. □

Another simple property of summations will be useful:

Lemma 7.150. Let $n \in \mathbb{N}$. Let u and v be two elements of $[n]$.
(a) We have $1 = [u < v] + [v \leq u]$.
(b) We have $\sum_{k \in [n]} [k > u] [k < v] = [u < v] (v - u - 1)$.
(c) We have $\sum_{k \in [n]} [k > u] [k \leq v] = [u < v] (v - u)$.

Proof of Lemma 7.150. **(a)** The logical statements $(v \leq u)$ and $(\text{not } u < v)$ are equivalent (since $(v \leq u) \iff (u \geq v) \iff (\text{not } u < v)$). Hence, Exercise 3.12 **(a)** (applied to $\mathcal{A} = (v \leq u)$ and $\mathcal{B} = (\text{not } u < v)$) shows that $[v \leq u] = [\text{not } u < v] = 1 - [u < v]$ (by Exercise 3.12 **(b)** (applied to $\mathcal{A} = (u < v)$)). Adding $[u < v]$ to both sides of this equality, we obtain $[u < v] + [v \leq u] = 1$. This proves Lemma 7.150 **(a)**.

(b) We have $u \in [n]$, so that $1 \leq u \leq n$. Similarly, $1 \leq v \leq n$.

Now,

$$\underbrace{\sum_{k \in [n]}}_{= \sum_{k=1}^{n}} [k > u] [k < v]$$
$$= \sum_{k=1}^{n} [k > u] [k < v] = \sum_{k=1}^{u} \underbrace{[k > u]}_{\substack{=0 \\ \text{(since we don't have } k>u \\ \text{(because } k \leq u))}} [k < v] + \sum_{k=u+1}^{n} \underbrace{[k > u]}_{\substack{=1 \\ \text{(since } k \geq u+1 > u)}} [k < v]$$
$$\left(\text{here, we have split the sum at } k = u, \text{ since } 1 \leq u \leq n\right)$$
$$= \underbrace{\sum_{k=1}^{u} 0 [k < v]}_{=0} + \sum_{k=u+1}^{n} [k < v] = \sum_{k=u+1}^{n} [k < v]. \tag{971}$$

We are in one of the following two cases:

Case 1: We have $u < v$.

Case 2: We don't have $u < v$.

Let us first consider Case 1. In this case, we have $u < v$. Thus, $u \leq v - 1$ (since u

and v are integers), so that $u+1 \leq v$. Now, (971) becomes

$$\begin{aligned}
&\sum_{k\in[n]} [k>u]\,[k<v] \\
&= \sum_{k=u+1}^{n} [k<v] = \sum_{k=u+1}^{v-1} \underbrace{[k<v]}_{\substack{=1\\ (\text{since } k\leq v-1<v)}} + \sum_{k=v}^{n} \underbrace{[k<v]}_{\substack{=0\\ (\text{since we don't have } k<v\\ (\text{because } k\geq v))}} \\
&\qquad (\text{here, we have split the sum at } k=v-1, \text{ since } u+1\leq v\leq n) \\
&= \sum_{k=u+1}^{v-1} 1 + \underbrace{\sum_{k=v}^{n} 0}_{=0} = \sum_{k=u+1}^{v-1} 1 = ((v-1)-u)\cdot 1 \qquad (\text{since } u\leq v-1) \\
&= v-u-1.
\end{aligned}$$

Comparing this with

$$\underbrace{[u<v]}_{\substack{=1\\ (\text{since } u<v)}} (v-u-1) = v-u-1,$$

we obtain $\sum_{k\in[n]} [k>u]\,[k<v] = [u<v]\,(v-u-1)$. Thus, Lemma 7.150 **(b)** is proven in Case 1.

Let us next consider Case 2. In this case, we don't have $u<v$. Thus, we have $u\geq v$. Now, (971) becomes

$$\sum_{k\in[n]} [k>u]\,[k<v] = \sum_{k=u+1}^{n} \underbrace{[k<v]}_{\substack{=0\\ (\text{since we don't have } k<v\\ (\text{because } k\geq u+1\geq u\geq v))}} = \sum_{k=u+1}^{n} 0 = 0.$$

Comparing this with

$$\underbrace{[u<v]}_{\substack{=0\\ (\text{since we don't have } u<v)}} (v-u-1) = 0\,(v-u-1) = 0,$$

we obtain $\sum_{k\in[n]} [k>u]\,[k<v] = [u<v]\,(v-u-1)$. Thus, Lemma 7.150 **(b)** is proven in Case 2.

We have now proven Lemma 7.150 **(b)** in each of the two Cases 1 and 2. Hence, Lemma 7.150 **(b)** always holds.

(c) We have $u\in[n]$, so that $1\leq u\leq n$. Similarly, $1\leq v\leq n$.

Now,

$$\sum_{k\in[n]} [k>u]\,[k\leq v] = \sum_{k=u+1}^{n} [k\leq v]. \tag{972}$$

(This can be proven in the exact same way as (971), except that all the "$[k < v]$" terms must be replaced by "$[k \leq v]$".)

We are in one of the following two cases:

Case 1: We have $u < v$.

Case 2: We don't have $u < v$.

Let us first consider Case 1. In this case, we have $u < v$. Thus, $u \leq v - 1$ (since u and v are integers), so that $u + 1 \leq v$. Now, (972) becomes

$$\begin{aligned}
&\sum_{k\in[n]} [k > u] [k \leq v] \\
&= \sum_{k=u+1}^{n} [k \leq v] = \sum_{k=u+1}^{v} \underbrace{[k \leq v]}_{\substack{=1 \\ (\text{since } k \leq v)}} + \sum_{k=v+1}^{n} \underbrace{[k \leq v]}_{\substack{=0 \\ (\text{since we don't have } k \leq v \\ (\text{because } k \geq v+1 > v))}} \\
&\qquad \left(\text{here, we have split the sum at } k = v\text{, since } u + 1 \leq v \leq n\right) \\
&= \sum_{k=u+1}^{v} 1 + \underbrace{\sum_{k=v+1}^{n} 0}_{=0} = \sum_{k=u+1}^{v} 1 = (v - u) \cdot 1 \qquad (\text{since } u \leq v - 1 \leq v) \\
&= v - u.
\end{aligned}$$

Comparing this with $\underbrace{[u < v]}_{\substack{=1 \\ (\text{since } u<v)}} (v - u) = v - u$, we obtain $\sum_{k\in[n]} [k > u] [k \leq v] = [u < v] (v - u)$. Thus, Lemma 7.150 **(c)** is proven in Case 1.

Let us next consider Case 2. In this case, we don't have $u < v$. Thus, we have $u \geq v$. Now, (972) becomes

$$\sum_{k\in[n]} [k > u] [k \leq v] = \sum_{k=u+1}^{n} \underbrace{[k \leq v]}_{\substack{=0 \\ (\text{since we don't have } k \leq v \\ (\text{because } k \geq u+1 > u \geq v))}} = \sum_{k=u+1}^{n} 0 = 0.$$

Comparing this with $\underbrace{[u < v]}_{\substack{=0 \\ (\text{since we don't have } u<v)}} (v - u) = 0 (v - u) = 0$, we obtain $\sum_{k\in[n]} [k > u] [k \leq v] = [u < v] (v - u)$. Thus, Lemma 7.150 **(c)** is proven in Case 2.

We have now proven Lemma 7.150 **(c)** in each of the two Cases 1 and 2. Hence, Lemma 7.150 **(c)** always holds. □

Solution to Exercise 5.28. We start with some simple observations.

Recall that S_n is the set of all permutations of $\{1, 2, \ldots, n\}$. In other words, S_n is the set of all permutations of $[n]$ (since $\{1, 2, \ldots, n\} = [n]$). Hence, σ is a permutation of $[n]$ (since $\sigma \in S_n$). In other words, σ is a bijective map $[n] \to [n]$.

Each $i \in [n]$ satisfies exactly one of the three statements $\sigma(i) < i$, $\sigma(i) = i$ and $\sigma(i) > i$. Hence, we can subdivide the sum $\sum_{i \in [n]} (\sigma(i) - i)$ as follows:

$$\begin{aligned}
\sum_{i \in [n]} (\sigma(i) - i) &= \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} \underbrace{(\sigma(i) - i)}_{= -(i - \sigma(i))} + \sum_{\substack{i \in [n]; \\ \sigma(i) = i}} \underbrace{(\sigma(i) - i)}_{\substack{=0 \\ (\text{since } \sigma(i) = i)}} + \sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) \\
&= \underbrace{\sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (-(i - \sigma(i)))}_{= - \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i))} + \underbrace{\sum_{\substack{i \in [n]; \\ \sigma(i) = i}} 0}_{=0} + \sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) \\
&= - \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i)) + \sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) \\
&= \sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) - \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i)).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) - \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i)) \\
&= \sum_{i \in [n]} (\sigma(i) - i) = \underbrace{\sum_{i \in [n]} \sigma(i)}_{\substack{= \sum_{i \in [n]} i \\ \text{(here, we have substituted } i \\ \text{for } \sigma(i) \text{ in the sum, since } \sigma \\ \text{is a bijection } [n] \to [n])}} - \sum_{i \in [n]} i = \sum_{i \in [n]} i - \sum_{i \in [n]} i = 0.
\end{aligned}$$

In other words,

$$\sum_{\substack{i \in [n]; \\ \sigma(i) > i}} (\sigma(i) - i) = \sum_{\substack{i \in [n]; \\ \sigma(i) < i}} (i - \sigma(i)). \tag{973}$$

(a) Each $i \in [n]$ satisfies exactly one of the three statements $\sigma(i) < i$, $\sigma(i) = i$

and $\sigma(i) > i$. Hence, we can subdivide the sum $\sum_{i \in [n]} |\sigma(i) - i|$ as follows:

$$\begin{aligned}
\sum_{i\in[n]} |\sigma(i) - i| &= \sum_{\substack{i\in[n];\\ \sigma(i)<i}} \underbrace{|\sigma(i) - i|}_{\substack{=-(\sigma(i)-i)\\ (\text{since } \sigma(i)-i<0\\ (\text{because } \sigma(i)<i))}} + \sum_{\substack{i\in[n];\\ \sigma(i)=i}} \left| \underbrace{\sigma(i) - i}_{\substack{=0\\ (\text{since } \sigma(i)=i)}} \right| + \sum_{\substack{i\in[n];\\ \sigma(i)>i}} \underbrace{|\sigma(i) - i|}_{\substack{=\sigma(i)-i\\ (\text{since } \sigma(i)-i>0\\ (\text{because } \sigma(i)>i))}} \\
&= \sum_{\substack{i\in[n];\\ \sigma(i)<i}} \underbrace{(-(\sigma(i) - i))}_{=i-\sigma(i)} + \sum_{\substack{i\in[n];\\ \sigma(i)=i}} \underbrace{|0|}_{=0} + \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i) \\
&= \underbrace{\sum_{\substack{i\in[n];\\ \sigma(i)<i}} (i - \sigma(i))}_{\substack{= \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i)-i)\\ (\text{by (973)})}} + \underbrace{\sum_{\substack{i\in[n];\\ \sigma(i)=i}} 0}_{=0} + \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i) \\
&= \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i) + \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i) = 2 \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i).
\end{aligned}$$

Now, the definition of $h(\sigma)$ yields

$$h(\sigma) = \sum_{i\in[n]} |\sigma(i) - i| = 2 \underbrace{\sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i)}_{\substack{= \sum_{\substack{i\in[n];\\ \sigma(i)<i}} (i-\sigma(i))\\ (\text{by (973)})}} = 2 \sum_{\substack{i\in[n];\\ \sigma(i)<i}} (i - \sigma(i)).$$

This solves Exercise 5.28 **(a)**.

(b) Let $\tau \in S_n$.

We have proven that σ is a bijection $[n] \to [n]$. Similarly, τ is a bijection $[n] \to [n]$.

The definition of $h(\tau)$ yields

$$h(\tau) = \sum_{i\in[n]} |\tau(i) - i|. \tag{974}$$

The definition of $h(\sigma \circ \tau)$ yields

$$h(\sigma \circ \tau) = \sum_{i\in[n]} \left| \underbrace{(\sigma \circ \tau)(i)}_{=\sigma(\tau(i))} - i \right| = \sum_{i\in[n]} |\sigma(\tau(i)) - i|. \tag{975}$$

But the definition of $h(\sigma)$ yields

$$h(\sigma) = \sum_{i\in[n]} |\sigma(i) - i| = \sum_{i\in[n]} |\sigma(\tau(i)) - \tau(i)|$$

(here, we have substituted $\tau(i)$ for i in the sum, since τ is a bijection $[n] \to [n]$). Adding (974) to this equality, we obtain

$$\begin{aligned} h(\sigma) + h(\tau) &= \sum_{i\in[n]} |\sigma(\tau(i)) - \tau(i)| + \sum_{i\in[n]} |\tau(i) - i| \\ &= \sum_{i\in[n]} \underbrace{(|\sigma(\tau(i)) - \tau(i)| + |\tau(i) - i|)}_{\substack{\geq |(\sigma(\tau(i))-\tau(i))+(\tau(i)-i)| \\ \text{(by Theorem 7.149} \\ \text{(applied to } x=\sigma(\tau(i))-\tau(i) \text{ and } y=\tau(i)-i))}} \\ &\geq \sum_{i\in[n]} \left| \underbrace{(\sigma(\tau(i)) - \tau(i)) + (\tau(i) - i)}_{=\sigma(\tau(i))-i} \right| = \sum_{i\in[n]} |\sigma(\tau(i)) - i| = h(\sigma\circ\tau) \end{aligned}$$

(by (975)). In other words, $h(\sigma\circ\tau) \leq h(\sigma) + h(\tau)$. This solves Exercise 5.28 **(b)**.

(d) Exercise 5.28 **(a)** yields $h(\sigma) = 2 \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i) - i)$. Dividing this equality by 2, we obtain

$$h(\sigma)/2 = \sum_{\substack{i\in[n];\\ \sigma(i)>i}} \left(\underbrace{\sigma(i)}_{\substack{\leq i+\ell_i(\sigma) \\ \text{(by Lemma 5.48 (c))}}} - i \right) \tag{976}$$

$$\leq \sum_{\substack{i\in[n];\\ \sigma(i)>i}} \underbrace{(i + \ell_i(\sigma) - i)}_{=\ell_i(\sigma)} = \sum_{\substack{i\in[n];\\ \sigma(i)>i}} \ell_i(\sigma). \tag{977}$$

Also, Exercise 5.28 **(a)** yields $h(\sigma) = 2 \sum_{\substack{i\in[n];\\ \sigma(i)<i}} (i - \sigma(i))$. Dividing this equality by 2, we obtain

$$h(\sigma)/2 = \sum_{\substack{i\in[n];\\ \sigma(i)<i}} (i - \sigma(i)). \tag{978}$$

On the other hand, each $i \in [n]$ satisfies

$$\ell_i(\sigma) \geq 0. \tag{979}$$

(This follows immediately from the definition of $\ell_i(\sigma)$.)

Proposition 5.46 yields

$$\ell(\sigma) = \ell_1(\sigma) + \ell_2(\sigma) + \cdots + \ell_n(\sigma) = \underbrace{\sum_{i=1}^{n}}_{\substack{= \sum_{i \in \{1,2,\ldots,n\}} = \sum_{i \in [n]} \\ (\text{since } \{1,2,\ldots,n\} = [n])}} \ell_i(\sigma)$$

$$= \sum_{i \in [n]} \ell_i(\sigma) \tag{980}$$

$$= \sum_{\substack{i \in [n]; \\ \sigma(i) \leq i}} \underbrace{\ell_i(\sigma)}_{\substack{\geq 0 \\ (\text{by (979)})}} + \underbrace{\sum_{\substack{i \in [n]; \\ \sigma(i) > i}} \ell_i(\sigma)}_{\substack{\geq h(\sigma)/2 \\ (\text{by (977)})}}$$

$$\left(\begin{array}{c} \text{since each } i \in [n] \text{ satisfies either } \sigma(i) \leq i \text{ or } \sigma(i) > i, \\ \text{but not both at the same time} \end{array} \right)$$

$$\geq \underbrace{\sum_{\substack{i \in [n]; \\ \sigma(i) \leq i}} 0}_{=0} + h(\sigma)/2 \geq h(\sigma)/2.$$

In other words, $h(\sigma)/2 \leq \ell(\sigma)$.

It now remains to prove that $\ell(\sigma) \leq h(\sigma)$. Our proof of this inequality shall follow the idea given in the Hint, but not literally: Instead of counting inversions, we will be summing truth values (which boils down to the same thing, but looks slicker on paper).

The equality (980) becomes

$$\ell(\sigma) = \sum_{i \in [n]} \underbrace{\ell_i(\sigma)}_{\substack{= \sum_{j \in [n]} [i<j][\sigma(i)>\sigma(j)] \\ (\text{by Lemma 7.113 } \textbf{(b)})}}$$

$$= \sum_{i \in [n]} \sum_{j \in [n]} [i < j] [\sigma(i) > \sigma(j)]. \tag{981}$$

Next, we claim:

Claim 1: Let $i \in [n]$ and $j \in [n]$. Then,

$$[i < j] [\sigma(i) > \sigma(j)] \leq [\sigma(i) > \sigma(j)] [\sigma(i) < j] + [j > i] [j \leq \sigma(i)].$$

[*Proof of Claim 1:* The logical statements $(i < j)$ and $(j > i)$ are equivalent. Hence, Exercise 3.12 **(a)** yields $[i < j] = [j > i]$.

We have $\sigma(i) \in [n]$ (since σ is a map $[n] \to [n]$). Now,

$$\begin{aligned}
& [i<j]\,[\sigma(i)>\sigma(j)] \\
&= [i<j]\,[\sigma(i)>\sigma(j)] \cdot \underbrace{1}_{\substack{=[\sigma(i)<j]+[j\leq\sigma(i)] \\ \text{(by Lemma 7.150 (a)} \\ \text{(applied to } u=\sigma(i) \text{ and } v=j))}} \\
&= [i<j]\,[\sigma(i)>\sigma(j)]\,([\sigma(i)<j]+[j\leq\sigma(i)]) \\
&= \underbrace{[i<j]}_{\substack{\leq 1 \\ \text{(since } [\mathcal{A}]\leq 1 \text{ for} \\ \text{any statement } \mathcal{A})}} [\sigma(i)>\sigma(j)]\,[\sigma(i)<j] + \underbrace{[i<j]}_{=[j>i]} \underbrace{[\sigma(i)>\sigma(j)]}_{\substack{\leq 1 \\ \text{(since } [\mathcal{A}]\leq 1 \text{ for} \\ \text{any statement } \mathcal{A})}} [j\leq\sigma(i)] \\
&\leq [\sigma(i)>\sigma(j)]\,[\sigma(i)<j] + [j>i]\,[j\leq\sigma(i)].
\end{aligned}$$

(Here, we have made tacit use of the fact that all terms in our computation are nonnegative[443]. Indeed, this fact allowed us to multiply inequalities.) Thus, Claim 1 is proven.]

Now, (981) becomes

$$\begin{aligned}
\ell(\sigma) &= \sum_{i\in[n]} \sum_{j\in[n]} \underbrace{[i<j]\,[\sigma(i)>\sigma(j)]}_{\substack{\leq[\sigma(i)>\sigma(j)][\sigma(i)<j]+[j>i][j\leq\sigma(i)] \\ \text{(by Claim 1)}}} \\
&\leq \sum_{i\in[n]} \sum_{j\in[n]} ([\sigma(i)>\sigma(j)]\,[\sigma(i)<j] + [j>i]\,[j\leq\sigma(i)]) \\
&= \underbrace{\sum_{i\in[n]} \sum_{j\in[n]}}_{=\sum_{j\in[n]}\sum_{i\in[n]}} [\sigma(i)>\sigma(j)]\,[\sigma(i)<j] + \sum_{i\in[n]} \sum_{j\in[n]} [j>i]\,[j\leq\sigma(i)] \\
&= \sum_{j\in[n]} \underbrace{\sum_{i\in[n]} [\sigma(i)>\sigma(j)]\,[\sigma(i)<j]}_{\substack{=\sum_{k\in[n]}[k>\sigma(j)][k<j] \\ \text{(here, we have substituted } k \text{ for } \sigma(i) \text{ in} \\ \text{the sum, since the map } \sigma:[n]\to[n] \text{ is a bijection)}}} + \sum_{i\in[n]} \underbrace{\sum_{j\in[n]} [j>i]\,[j\leq\sigma(i)]}_{\substack{=\sum_{k\in[n]}[k>i][k\leq\sigma(i)] \\ \text{(here, we have renamed the summation} \\ \text{index } j \text{ as } k)}} \\
&= \underbrace{\sum_{j\in[n]} \sum_{k\in[n]} [k>\sigma(j)]\,[k<j]}_{\substack{=\sum_{i\in[n]}\sum_{k\in[n]}[k>\sigma(i)][k<i] \\ \text{(here, we have renamed the summation index } j \\ \text{as } i \text{ in the outer sum)}}} + \sum_{i\in[n]} \sum_{k\in[n]} [k>i]\,[k\leq\sigma(i)] \\
&= \sum_{i\in[n]} \sum_{k\in[n]} [k>\sigma(i)]\,[k<i] + \sum_{i\in[n]} \sum_{k\in[n]} [k>i]\,[k\leq\sigma(i)]. \qquad (982)
\end{aligned}$$

Now, fix $i \in [n]$. Then, $\sigma(i) \in [n]$ (since σ is a map $[n] \to [n]$). Hence, Lemma

[443]because $[\mathcal{A}] \geq 0$ for any statement $\mathcal{A}$

7.150 **(b)** (applied to $u = \sigma(i)$ and $v = i$) yields

$$\sum_{k\in[n]} [k > \sigma(i)] [k < i] = [\sigma(i) < i] (i - \sigma(i) - 1). \qquad (983)$$

Moreover, Lemma 7.150 **(c)** (applied to $u = i$ and $v = \sigma(i)$) yields

$$\sum_{k\in[n]} [k > i] [k \leq \sigma(i)] = [i < \sigma(i)] (\sigma(i) - i). \qquad (984)$$

Now, forget that we have fixed i. We thus have proven (983) and (984) for each $i \in [n]$.

Now,

$$\begin{aligned}
&\sum_{i\in[n]} \underbrace{\sum_{k\in[n]} [k > \sigma(i)] [k < i]}_{\substack{=[\sigma(i)<i](i-\sigma(i)-1) \\ \text{(by (983))}}} \\
&= \sum_{i\in[n]} [\sigma(i) < i] (i - \sigma(i) - 1) \\
&= \sum_{\substack{i\in[n]; \\ \sigma(i)<i}} \underbrace{[\sigma(i) < i]}_{\substack{=1 \\ \text{(since } \sigma(i)<i)}} (i - \sigma(i) - 1) + \sum_{\substack{i\in[n]; \\ \text{not } \sigma(i)<i}} \underbrace{[\sigma(i) < i]}_{\substack{=0 \\ \text{(since we don't have } \sigma(i)<i)}} (i - \sigma(i) - 1) \\
&\qquad \left(\begin{array}{c} \text{since each } i \in [n] \text{ satisfies either } \sigma(i) < i \text{ or } (\text{not } \sigma(i) < i), \\ \text{but not both at the same time} \end{array} \right) \\
&= \sum_{\substack{i\in[n]; \\ \sigma(i)<i}} (i - \sigma(i) - 1) + \underbrace{\sum_{\substack{i\in[n]; \\ \text{not } \sigma(i)<i}} 0 (i - \sigma(i) - 1)}_{=0} = \sum_{\substack{i\in[n]; \\ \sigma(i)<i}} \underbrace{(i - \sigma(i) - 1)}_{\leq i-\sigma(i)} \\
&\leq \sum_{\substack{i\in[n]; \\ \sigma(i)<i}} (i - \sigma(i)) = h(\sigma)/2 \qquad \text{(by (978))}.
\end{aligned} \qquad (985)$$

Furthermore,

$$\sum_{i\in[n]} \underbrace{\sum_{k\in[n]} [k>i]\,[k\leq\sigma(i)]}_{\substack{=[i<\sigma(i)](\sigma(i)-i)\\ \text{(by (984))}}}$$
$$= \sum_{i\in[n]} [i<\sigma(i)]\,(\sigma(i)-i)$$
$$= \sum_{\substack{i\in[n];\\ i<\sigma(i)}} \underbrace{[i<\sigma(i)]}_{\substack{=1\\ \text{(since } i<\sigma(i))}} (\sigma(i)-i) + \sum_{\substack{i\in[n];\\ \text{not } i<\sigma(i)}} \underbrace{[i<\sigma(i)]}_{\substack{=0\\ \text{(since we don't have } i<\sigma(i))}} (\sigma(i)-i)$$
$$\left(\begin{array}{c}\text{since each } i\in[n] \text{ satisfies either } i<\sigma(i) \text{ or } (\text{not } i<\sigma(i)),\\ \text{but not both at the same time}\end{array}\right)$$
$$= \sum_{\substack{i\in[n];\\ i<\sigma(i)}} (\sigma(i)-i) + \underbrace{\sum_{\substack{i\in[n];\\ \text{not } i<\sigma(i)}} 0\,(\sigma(i)-i)}_{=0} = \underbrace{\sum_{\substack{i\in[n];\\ i<\sigma(i)}}}_{\substack{=\sum_{\substack{i\in[n];\\ \sigma(i)>i}}\\ \text{(since the condition } (i<\sigma(i))\\ \text{is equivalent to } (\sigma(i)>i))}} (\sigma(i)-i)$$
$$= \sum_{\substack{i\in[n];\\ \sigma(i)>i}} (\sigma(i)-i) = h(\sigma)/2 \qquad (\text{by (976)}). \tag{986}$$

Now, (982) becomes

$$\ell(\sigma) \leq \underbrace{\sum_{i\in[n]}\sum_{k\in[n]} [k>\sigma(i)]\,[k<i]}_{\substack{\leq h(\sigma)/2\\ \text{(by (985))}}} + \underbrace{\sum_{i\in[n]}\sum_{k\in[n]} [k>i]\,[k\leq\sigma(i)]}_{\substack{=h(\sigma)/2\\ \text{(by (986))}}}$$
$$\leq h(\sigma)/2 + h(\sigma)/2 = h(\sigma).$$

Thus, $\ell(\sigma) \leq h(\sigma)$ is proven.

We have now proven both $h(\sigma)/2 \leq \ell(\sigma)$ and $\ell(\sigma) \leq h(\sigma)$. Thus, $h(\sigma)/2 \leq \ell(\sigma) \leq h(\sigma)$ follows. This solves Exercise 5.28 **(d)**.

(c) Let $k \in \{1,2,\ldots,n-1\}$. We have $\ell(s_k) = 1$. (This was shown during our proof of Proposition 5.15 **(b)**.) But Exercise 5.28 **(d)** (applied to s_k instead of σ) yields $h(s_k)/2 \leq \ell(s_k) \leq h(s_k)$. From $h(s_k)/2 \leq \ell(s_k)$, we obtain $h(s_k) \leq 2\underbrace{\ell(s_k)}_{=1} = 2$. (It is not hard to check that $h(s_k) = 2$, but we will not need this fact.)

Now, Exercise 5.28 **(b)** (applied to s_k and σ instead of σ and τ) yields $h(s_k\circ\sigma) \leq \underbrace{h(s_k)}_{\leq 2} + h(\sigma) \leq 2 + h(\sigma) = h(\sigma) + 2$. This solves Exercise 5.28 **(c)**. $\square$

7.67. Solution to Exercise 5.29

We shall now prepare for the solution of Exercise 5.29. First, we introduce a notation:

Definition 7.151. If X, X', Y and Y' are four sets and if $\alpha : X \to X'$ and $\beta : Y \to Y'$ are two maps, then $\alpha \times \beta$ will denote the map

$$X \times Y \to X' \times Y',$$
$$(x, y) \mapsto (\alpha(x), \beta(y)).$$

Lemma 7.152. Let $n \in \mathbb{N}$. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Let $\alpha \in S_n$ and $\beta \in S_n$. Then, the map $\alpha \times \beta : [n] \times [n] \to [n] \times [n]$ is invertible, and its inverse is $\alpha^{-1} \times \beta^{-1}$.

Proof of Lemma 7.152. Recall that S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$. In other words, S_n is the set of all permutations of the set $[n]$ (since $\{1, 2, \ldots, n\} = [n]$).

We know that $\alpha \in S_n$. In other words, α is a permutation of the set $[n]$ (since S_n is the set of all permutations of the set $[n]$). In other words, α is a bijective map $[n] \to [n]$. Similarly, β is a bijective map $[n] \to [n]$. Hence, Definition 7.151 defines a map $\alpha \times \beta : [n] \times [n] \to [n] \times [n]$.

The map α is bijective, and thus invertible. Hence, its inverse map $\alpha^{-1} : [n] \to [n]$ is well-defined. Similarly, $\beta^{-1} : [n] \to [n]$ is well-defined. Thus, Definition 7.151 defines a map $\alpha^{-1} \times \beta^{-1} : [n] \times [n] \to [n] \times [n]$.

Now, every $(i, j) \in [n] \times [n]$ satisfies

$$\begin{aligned}
&\left(\left(\alpha^{-1} \times \beta^{-1}\right) \circ (\alpha \times \beta)\right)(i, j) \\
&= \left(\alpha^{-1} \times \beta^{-1}\right) \left(\underbrace{(\alpha \times \beta)(i, j)}_{\substack{=(\alpha(i), \beta(j)) \\ \text{(by the definition of } \alpha \times \beta)}} \right) \\
&= \left(\alpha^{-1} \times \beta^{-1}\right)(\alpha(i), \beta(j)) = \left(\underbrace{\alpha^{-1}(\alpha(i))}_{=i}, \underbrace{\beta^{-1}(\beta(j))}_{=j} \right) \\
&\qquad \left(\text{by the definition of } \alpha^{-1} \times \beta^{-1}\right) \\
&= (i, j) = \operatorname{id}(i, j).
\end{aligned}$$

Thus, $\left(\alpha^{-1} \times \beta^{-1}\right) \circ (\alpha \times \beta) = \operatorname{id}$. Similarly, $(\alpha \times \beta) \circ \left(\alpha^{-1} \times \beta^{-1}\right) = \operatorname{id}$. These two equalities show that the maps $\alpha \times \beta$ and $\alpha^{-1} \times \beta^{-1}$ are mutually inverse. Hence, the map $\alpha \times \beta : [n] \times [n] \to [n] \times [n]$ is invertible, and its inverse is $\alpha^{-1} \times \beta^{-1}$. This proves Lemma 7.152. $\square$

Lemma 7.153. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ and $\tau \in S_n$. Then:

(a) We have $\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau \subseteq (\tau \times \tau)^{-1} (\operatorname{Inv} \sigma)$. (Here, $\tau \times \tau$ is defined as in Definition 7.151.)

(b) We have $|\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau| \leq |\operatorname{Inv} \sigma|$.

Proof of Lemma 7.153. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Lemma 7.152 (applied to $\alpha = \tau$ and $\beta = \tau$) yields that the map $\tau \times \tau : [n] \times [n] \to [n] \times [n]$ is invertible, and its inverse is $\tau^{-1} \times \tau^{-1}$. Thus, $(\tau \times \tau)^{-1} = \tau^{-1} \times \tau^{-1}$.

(a) Let $c \in \operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau$. Thus, $c \in \operatorname{Inv}(\sigma \circ \tau)$ but $c \notin \operatorname{Inv} \tau$.

We have $c \in \operatorname{Inv}(\sigma \circ \tau)$. In other words, c is an inversion of $\sigma \circ \tau$ (since $\operatorname{Inv}(\sigma \circ \tau)$ is the set of all inversions of $\sigma \circ \tau$). In other words, c is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $(\sigma \circ \tau)(i) > (\sigma \circ \tau)(j)$ (by the definition of an "inversion of $\sigma \circ \tau$"). In other words, there exists a pair (i, j) of integers satisfying $1 \leq i < j \leq n$, $(\sigma \circ \tau)(i) > (\sigma \circ \tau)(j)$ and $c = (i, j)$. Let us denote this pair (i, j) by (u, v). Thus, (u, v) is a pair of integers satisfying $1 \leq u < v \leq n$, $(\sigma \circ \tau)(u) > (\sigma \circ \tau)(v)$ and $c = (u, v)$.

But $\tau(u) < \tau(v)$ [444]. Also, $1 \leq \tau(u)$ (since $\tau(u) \in \{1, 2, \ldots, n\}$) and $\tau(v) \leq n$ (since $\tau(v) \in \{1, 2, \ldots, n\}$). Finally, $\sigma(\tau(u)) = (\sigma \circ \tau)(u) > (\sigma \circ \tau)(v) = \sigma(\tau(v))$.

Thus, $(\tau(u), \tau(v))$ is a pair of integers satisfying $1 \leq \tau(u) < \tau(v) \leq n$ and $\sigma(\tau(u)) > \sigma(\tau(v))$. In other words, $(\tau(u), \tau(v))$ is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. In other words, $(\tau(u), \tau(v))$ is an inversion of σ (by the definition of an "inversion of σ"). In other words, $(\tau(u), \tau(v)) \in \operatorname{Inv} \sigma$ (since $\operatorname{Inv} \sigma$ is the set of all inversions of σ).

Now, $c = (u, v) \in [n] \times [n]$ (since both u and v belong to $[n]$), and we have

$$(\tau \times \tau)\left(\underbrace{c}_{=(u,v)}\right) = (\tau \times \tau)(u, v) = (\tau(u), \tau(v)) \qquad \text{(by the definition of } \tau \times \tau\text{)}$$
$$\in \operatorname{Inv} \sigma,$$

so that $c \in (\tau \times \tau)^{-1} (\operatorname{Inv} \sigma)$.

Now, forget that we fixed c. We thus have shown that $c \in (\tau \times \tau)^{-1} (\operatorname{Inv} \sigma)$ for every $c \in \operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau$. In other words, $\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau \subseteq (\tau \times \tau)^{-1} (\operatorname{Inv} \sigma)$. This proves Lemma 7.153 **(a)**.

---

[444] *Proof.* Assume the contrary. Thus, $\tau(u) \geq \tau(v)$. But $\tau \in S_n$. In other words, τ is a permutation of the set $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$). Thus, the map τ is bijective, and therefore also injective. But $u < v$, so that $u \neq v$ and therefore $\tau(u) \neq \tau(v)$ (since τ is injective). Combining this with $\tau(u) \geq \tau(v)$, we obtain $\tau(u) > \tau(v)$.

Now, we know that (u, v) is a pair of integers and satisfies $1 \leq u < v \leq n$ and $\tau(u) > \tau(v)$. In other words, (u, v) is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\tau(i) > \tau(j)$. In other words, (u, v) is an inversion of τ (by the definition of an "inversion of τ"). In other words, $(u, v) \in \operatorname{Inv} \tau$ (since $\operatorname{Inv} \tau$ is the set of all inversions of τ). But this contradicts $(u, v) = c \notin \operatorname{Inv} \tau$. This contradiction shows that our assumption was false, qed.

(b) Lemma 7.153 **(a)** yields $\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv}\tau \subseteq (\tau \times \tau)^{-1}(\operatorname{Inv}\sigma)$. Thus,

$$\left|\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv}\tau\right| \leq \left|(\tau \times \tau)^{-1}(\operatorname{Inv}\sigma)\right| = \left|\operatorname{Inv}\sigma\right|$$

(since $\tau \times \tau$ is a bijection (since the map $\tau \times \tau$ is invertible)). This proves Lemma 7.153 **(b)**. □

Lemma 7.154. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Then, $\ell(\sigma) = |\operatorname{Inv}\sigma|$.

Proof of Lemma 7.154. We have

$$\begin{aligned} \ell(\sigma) &= (\text{the number of inversions of } \sigma) \qquad (\text{by the definition of } \ell(\sigma)) \\ &= \left| \underbrace{(\text{the set of all inversions of } \sigma)}_{\substack{=\operatorname{Inv}(\sigma) \\ (\text{since } \operatorname{Inv}(\sigma) \text{ is the set of all inversions of } \sigma)}} \right| \\ &= |\operatorname{Inv}(\sigma)|. \end{aligned}$$

This proves Lemma 7.154. □

Before we step to the solution to Exercise 5.29, let us make a short digression and use Lemma 7.153 to give a new solution of Exercise 5.2 **(c)**:

*Third solution to Exercise 5.2 **(c)**.* Let $\sigma \in S_n$ and $\tau \in S_n$. We have $|A \setminus B| \geq |A| - |B|$ for any two finite sets A and B [445]. Applying this to $A = \operatorname{Inv}(\sigma \circ \tau)$ and $B = \operatorname{Inv}\tau$, we obtain $|\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv}\tau| \geq |\operatorname{Inv}(\sigma \circ \tau)| - |\operatorname{Inv}\tau|$. Thus,

$$\begin{aligned} |\operatorname{Inv}(\sigma \circ \tau)| - |\operatorname{Inv}\tau| \leq |\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv}\tau| \leq |\operatorname{Inv}\sigma| \qquad (\text{by Lemma 7.153 } \textbf{(b)}) \\ = \ell(\sigma) \qquad (\text{by Lemma 7.154}). \end{aligned}$$

Now,

$$\underbrace{\ell(\sigma \circ \tau)}_{\substack{=|\operatorname{Inv}(\sigma \circ \tau)| \\ (\text{by Lemma 7.154} \\ (\text{applied to } \sigma \circ \tau \text{ instead of } \sigma))}} - \underbrace{\ell(\tau)}_{\substack{=|\operatorname{Inv}\tau| \\ (\text{by Lemma 7.154} \\ (\text{applied to } \tau \text{ instead of } \sigma))}} = |\operatorname{Inv}(\sigma \circ \tau)| - |\operatorname{Inv}\tau| \leq \ell(\sigma).$$

In other words, $\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau)$. Thus, Exercise 5.2 **(c)** is solved again. □

[445] *Proof.* Let A and B be two finite sets. Then, $A \setminus B = A \setminus (A \cap B)$, so that

$$\begin{aligned} |A \setminus B| = |A \setminus (A \cap B)| = |A| - \underbrace{|A \cap B|}_{\substack{\leq |B| \\ (\text{since } A \cap B \subseteq B)}} \qquad (\text{since } A \cap B \subseteq A) \\ \geq |A| - |B|, \end{aligned}$$

qed.

Now, let us finally solve Exercise 5.29:

Solution to Exercise 5.29. **(a)** We first observe that

$$\underbrace{\ell(\sigma\circ\tau)}_{\substack{=|\operatorname{Inv}(\sigma\circ\tau)| \\ \text{(by Lemma 7.154} \\ \text{(applied to } \sigma\circ\tau \\ \text{instead of } \sigma))}} - \underbrace{\ell(\tau)}_{\substack{=|\operatorname{Inv}\tau| \\ \text{(by Lemma 7.154} \\ \text{(applied to } \tau \\ \text{instead of } \sigma))}} = |\operatorname{Inv}(\sigma\circ\tau)| - |\operatorname{Inv}\tau|. \tag{987}$$

Let us now prove the logical implication

$$(\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)) \implies (\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)). \tag{988}$$

[*Proof of (988):* Assume that $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$ holds. We will prove that $\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)$.

If two finite sets A and B satisfy $|A \setminus B| \leq |A| - |B|$, then

$$B \subseteq A \tag{989}$$

[446].

Now, Lemma 7.153 **(b)** yields

$$\begin{aligned} |\operatorname{Inv}(\sigma\circ\tau) \setminus \operatorname{Inv}\tau| &\leq |\operatorname{Inv}\sigma| = \ell(\sigma) && \text{(by Lemma 7.154)} \\ &= \ell(\sigma\circ\tau) - \ell(\tau) && (\text{since } \ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)) \\ &= |\operatorname{Inv}(\sigma\circ\tau)| - |\operatorname{Inv}\tau| && (\text{by (987)}). \end{aligned}$$

Thus, (989) (applied to $A = \operatorname{Inv}(\sigma\circ\tau)$ and $B = \operatorname{Inv}\tau$) yields $\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)$.

Now, forget our assumption that $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$. We thus have proven that if $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$, then $\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)$. In other words, we have proven the implication (988).]

Let us next prove the logical implication

$$(\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)) \implies (\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)). \tag{990}$$

[446] *Proof of (989):* Let A and B be two finite sets satisfying $|A \setminus B| \leq |A| - |B|$.

We have $A \setminus (A \cap B) = A \setminus B$, so that $|A \setminus (A \cap B)| = |A \setminus B| \leq |A| - |B|$. Adding $|B|$ to both sides of this inequality, we obtain $|A \setminus (A \cap B)| + |B| \leq |A|$. Hence,

$$|A| \geq \underbrace{|A \setminus (A \cap B)|}_{\substack{=|A|-|A\cap B| \\ (\text{since } A\cap B\subseteq A)}} + |B| = |A| - |A \cap B| + |B|.$$

Subtracting $|A|$ from both sides of this inequality, we obtain $0 \geq -|A \cap B| + |B|$. In other words, $|A \cap B| \geq |B|$. Also, clearly, $A \cap B$ is a subset of B.

But B is a finite set. Hence, the only subset of B having size $\geq |B|$ is B itself. In other words, if C is a subset of B satisfying $|C| \geq |B|$, then $C = B$. Applying this to $C = A \cap B$, we obtain $A \cap B = B$ (since $A \cap B$ is a subset of B satisfying $|A \cap B| \geq |B|$). Hence, $B = A \cap B \subseteq A$. This proves (989).

[*Proof of (990):* Assume that $\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)$ holds. We will prove that $\ell(\sigma\circ\tau) = \ell(\sigma) + \ell(\tau)$.

Consider the map $\tau\times\tau$ defined as in Definition 7.151. Let $[n]$ denote the set $\{1,2,\ldots,n\}$. Lemma 7.152 (applied to $\alpha = \tau$ and $\beta = \tau$) yields that the map $\tau\times\tau : [n]\times[n] \to [n]\times[n]$ is invertible, and its inverse is $\tau^{-1}\times\tau^{-1}$. Thus, $(\tau\times\tau)^{-1} = \tau^{-1}\times\tau^{-1}$.

Lemma 7.153 **(a)** shows that

$$\operatorname{Inv}(\sigma\circ\tau)\setminus\operatorname{Inv}\tau \subseteq (\tau\times\tau)^{-1}(\operatorname{Inv}\sigma). \tag{991}$$

We shall now prove the reverse inclusion, i.e., we shall prove that $(\tau\times\tau)^{-1}(\operatorname{Inv}\sigma) \subseteq \operatorname{Inv}(\sigma\circ\tau)\setminus\operatorname{Inv}\tau$.

Indeed, fix $c \in (\tau\times\tau)^{-1}(\operatorname{Inv}\sigma)$. Thus, $c\in[n]\times[n]$ and $(\tau\times\tau)(c)\in\operatorname{Inv}\sigma$.

We have $(\tau\times\tau)(c)\in\operatorname{Inv}\sigma$. In other words, $(\tau\times\tau)(c)$ is an inversion of σ (since $\operatorname{Inv}\sigma$ is the set of all inversions of σ). In other words, $(\tau\times\tau)(c)$ is a pair (i,j) of integers satisfying $1\leq i<j\leq n$ and $\sigma(i)>\sigma(j)$ (by the definition of an "inversion of σ"). In other words, there exists a pair (i,j) of integers satisfying $1\leq i<j\leq n$, $\sigma(i)>\sigma(j)$ and $(\tau\times\tau)(c)=(i,j)$. Let us denote this pair (i,j) by (u,v). Thus, (u,v) is a pair of integers satisfying $1\leq u<v\leq n$, $\sigma(u)>\sigma(v)$ and $(\tau\times\tau)(c)=(u,v)$.

From $(\tau\times\tau)(c)=(u,v)$, we obtain

$$\begin{aligned} c &= \underbrace{(\tau\times\tau)^{-1}}_{=\tau^{-1}\times\tau^{-1}}(u,v) \qquad \text{(since the map } \tau\times\tau \text{ is invertible)} \\ &= \left(\tau^{-1}\times\tau^{-1}\right)(u,v) = \left(\tau^{-1}(u),\tau^{-1}(v)\right) \end{aligned}$$

(by the definition of $\tau^{-1}\times\tau^{-1}$).

Notice that

$$(\sigma\circ\tau)\left(\tau^{-1}(v)\right) = \sigma\left(\underbrace{\tau\left(\tau^{-1}(v)\right)}_{=v}\right) = \sigma(v) \tag{992}$$

and

$$\begin{aligned} (\sigma\circ\tau)\left(\tau^{-1}(u)\right) &= \sigma\left(\underbrace{\tau\left(\tau^{-1}(u)\right)}_{=u}\right) = \sigma(u) \\ &> \sigma(v) = (\sigma\circ\tau)\left(\tau^{-1}(v)\right) \end{aligned} \tag{993}$$

(by (992)).

Let us now prove that $\tau^{-1}(u) < \tau^{-1}(v)$. Indeed, let us assume the contrary. Thus, $\tau^{-1}(u)\geq\tau^{-1}(v)$. But $u\neq v$ (since $u<v$), so that $\tau^{-1}(u)\neq\tau^{-1}(v)$. Combined with $\tau^{-1}(u)\geq\tau^{-1}(v)$, this yields $\tau^{-1}(u)>\tau^{-1}(v)$. In other words,

$\tau^{-1}(v) < \tau^{-1}(u)$. Also, $1 \leq \tau^{-1}(v)$ (since $\tau^{-1}(v) \in \{1,2,\ldots,n\}$) and $\tau^{-1}(u) \leq n$ (since $\tau^{-1}(u) \in \{1,2,\ldots,n\}$). Finally, $\tau\left(\tau^{-1}(v)\right) = v > u = \tau\left(\tau^{-1}(u)\right)$.

Thus, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is a pair of integers satisfying $1 \leq \tau^{-1}(v) < \tau^{-1}(u) \leq n$ and $\tau\left(\tau^{-1}(v)\right) > \tau\left(\tau^{-1}(u)\right)$. In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\tau(i) > \tau(j)$. In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is an inversion of τ (by the definition of an "inversion of τ"). In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right) \in \operatorname{Inv}\tau$ (since $\operatorname{Inv}\tau$ is the set of all inversions of τ). Hence, $\left(\tau^{-1}(v), \tau^{-1}(u)\right) \in \operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma\circ\tau)$. In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is an inversion of $\sigma\circ\tau$ (since $\operatorname{Inv}(\sigma\circ\tau)$ is the set of all inversions of $\sigma\circ\tau$). In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $(\sigma\circ\tau)(i) > (\sigma\circ\tau)(j)$ (by the definition of an "inversion of $\sigma\circ\tau$"). In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is a pair of integers satisfying $1 \leq \tau^{-1}(v) < \tau^{-1}(u) \leq n$ and $(\sigma\circ\tau)\left(\tau^{-1}(v)\right) > (\sigma\circ\tau)\left(\tau^{-1}(u)\right)$. But $(\sigma\circ\tau)\left(\tau^{-1}(v)\right) > (\sigma\circ\tau)\left(\tau^{-1}(u)\right) > (\sigma\circ\tau)\left(\tau^{-1}(v)\right)$ (by (993)), which is absurd. This contradiction proves that our assumption was wrong. Hence, $\tau^{-1}(u) < \tau^{-1}(v)$ is proven.

Now, $1 \leq \tau^{-1}(u)$ (since $\tau^{-1}(u) \in \{1,2,\ldots,n\}$) and $\tau^{-1}(v) \leq n$ (since $\tau^{-1}(v) \in \{1,2,\ldots,n\}$). Finally, recall that (993) holds. Thus, $\left(\tau^{-1}(u), \tau^{-1}(v)\right)$ is a pair of integers satisfying $1 \leq \tau^{-1}(u) < \tau^{-1}(v) \leq n$ and $(\sigma\circ\tau)\left(\tau^{-1}(u)\right) > (\sigma\circ\tau)\left(\tau^{-1}(v)\right)$. In other words, $\left(\tau^{-1}(u), \tau^{-1}(v)\right)$ is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $(\sigma\circ\tau)(i) > (\sigma\circ\tau)(j)$. In other words, $\left(\tau^{-1}(u), \tau^{-1}(v)\right)$ is an inversion of $\sigma\circ\tau$ (by the definition of an "inversion of $\sigma\circ\tau$"). In other words, $\left(\tau^{-1}(u), \tau^{-1}(v)\right) \in \operatorname{Inv}(\sigma\circ\tau)$ (since $\operatorname{Inv}(\sigma\circ\tau)$ is the set of all inversions of $\sigma\circ\tau$).

Hence, $c = \left(\tau^{-1}(u), \tau^{-1}(v)\right) \in \operatorname{Inv}(\sigma\circ\tau)$.

Let us now prove that $c \notin \operatorname{Inv}\tau$. Indeed, assume the contrary. Thus, $c \in \operatorname{Inv}\tau$. Thus, $\left(\tau^{-1}(u), \tau^{-1}(v)\right) = c \in \operatorname{Inv}\tau$. In other words, $\left(\tau^{-1}(u), \tau^{-1}(v)\right)$ is an inversion of τ (since $\operatorname{Inv}\tau$ is the set of all inversions of τ). In other words, $\left(\tau^{-1}(u), \tau^{-1}(v)\right)$ is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\tau(i) > \tau(j)$ (by the definition of an "inversion of τ"). In other words, $\left(\tau^{-1}(u), \tau^{-1}(v)\right)$ is a pair of integers satisfying $1 \leq \tau^{-1}(u) < \tau^{-1}(v) \leq n$ and $\tau\left(\tau^{-1}(u)\right) > \tau\left(\tau^{-1}(v)\right)$. But $\tau\left(\tau^{-1}(u)\right) > \tau\left(\tau^{-1}(v)\right) = v$ contradicts $\tau\left(\tau^{-1}(u)\right) = u < v$. This contradiction proves that our assumption was wrong. Hence, $c \notin \operatorname{Inv}\tau$ is proven.

Combining $c \in \operatorname{Inv}(\sigma\circ\tau)$ with $c \notin \operatorname{Inv}\tau$, we obtain $c \in \operatorname{Inv}(\sigma\circ\tau) \setminus \operatorname{Inv}\tau$.

Now, forget that we fixed c. We thus have shown that $c \in \operatorname{Inv}(\sigma\circ\tau) \setminus \operatorname{Inv}\tau$ for each $c \in (\tau\times\tau)^{-1}(\operatorname{Inv}\sigma)$. In other words, we have proven the inclusion

$$(\tau\times\tau)^{-1}(\operatorname{Inv}\sigma) \subseteq \operatorname{Inv}(\sigma\circ\tau) \setminus \operatorname{Inv}\tau.$$

Combining this with the inclusion (991), we obtain

$$\operatorname{Inv}(\sigma\circ\tau) \setminus \operatorname{Inv}\tau = (\tau\times\tau)^{-1}(\operatorname{Inv}\sigma).$$

Hence,

$$|\operatorname{Inv}(\sigma\circ\tau) \setminus \operatorname{Inv}\tau| = \left|(\tau\times\tau)^{-1}(\operatorname{Inv}\sigma)\right| = |\operatorname{Inv}\sigma|$$

(since the map $\tau \times \tau$ is a bijection (since $\tau \times \tau$ is invertible)). Thus,

$$|\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau| = |\operatorname{Inv} \sigma| = \ell(\sigma)$$

(by Lemma 7.154). Hence,

$$\begin{aligned}\ell(\sigma) &= |\operatorname{Inv}(\sigma \circ \tau) \setminus \operatorname{Inv} \tau| = |\operatorname{Inv}(\sigma \circ \tau)| - |\operatorname{Inv} \tau| \qquad (\text{since } \operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)) \\ &= \ell(\sigma \circ \tau) - \ell(\tau) \qquad (\text{by (987)}).\end{aligned}$$

In other words, $\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)$.

Now, forget our assumption that $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$. We thus have proven that if $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$, then $\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)$. In other words, we have proven the implication (990).]

Combining the two implications (988) and (990), we obtain the logical equivalence

$$(\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)) \iff (\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)).$$

In other words, $\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)$ holds if and only if $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$. This solves Exercise 5.29 **(a)**.

(b) Exercise 5.2 **(f)** (applied to $\sigma \circ \tau$ instead of σ) yields $\ell(\sigma \circ \tau) = \ell\left(\underbrace{(\sigma \circ \tau)^{-1}}_{=\tau^{-1} \circ \sigma^{-1}}\right) = \ell(\tau^{-1} \circ \sigma^{-1})$. Exercise 5.2 **(f)** (applied to τ instead of σ) yields $\ell(\tau) = \ell(\tau^{-1})$. Exercise 5.2 **(f)** yields $\ell(\sigma) = \ell(\sigma^{-1})$. Hence, $\underbrace{\ell(\sigma)}_{=\ell(\sigma^{-1})} + \underbrace{\ell(\tau)}_{=\ell(\tau^{-1})} = \ell(\sigma^{-1}) + \ell(\tau^{-1}) = \ell(\tau^{-1}) + \ell(\sigma^{-1})$.

Exercise 5.29 **(a)** (applied to τ^{-1} and σ^{-1} instead of σ and τ) yields that $\ell(\tau^{-1} \circ \sigma^{-1}) = \ell(\tau^{-1}) + \ell(\sigma^{-1})$ holds if and only if $\operatorname{Inv}(\sigma^{-1}) \subseteq \operatorname{Inv}(\tau^{-1} \circ \sigma^{-1})$. In light of the equalities $\ell(\tau^{-1} \circ \sigma^{-1}) = \ell(\sigma \circ \tau)$ and $\ell(\tau^{-1}) + \ell(\sigma^{-1}) = \ell(\sigma) + \ell(\tau)$, we can rewrite this as follows:

$$\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau) \text{ holds if and only if } \operatorname{Inv}\left(\sigma^{-1}\right) \subseteq \operatorname{Inv}\left(\tau^{-1} \circ \sigma^{-1}\right).$$

This solves Exercise 5.29 **(b)**.

(c) Exercise 5.29 **(a)** (applied to $\tau \circ \sigma^{-1}$ and σ instead of σ and τ) yields that $\ell(\tau \circ \sigma^{-1} \circ \sigma) = \ell(\tau \circ \sigma^{-1}) + \ell(\sigma)$ holds if and only if $\operatorname{Inv} \sigma \subseteq \operatorname{Inv}(\tau \circ \sigma^{-1} \circ \sigma)$. In light of $\tau \circ \underbrace{\sigma^{-1} \circ \sigma}_{=\operatorname{id}} = \tau \circ \operatorname{id} = \tau$, this rewrites as follows:

$$\ell(\tau) = \ell\left(\tau \circ \sigma^{-1}\right) + \ell(\sigma) \text{ holds if and only if } \operatorname{Inv} \sigma \subseteq \operatorname{Inv} \tau.$$

In other words, $\operatorname{Inv} \sigma \subseteq \operatorname{Inv} \tau$ holds if and only if $\ell(\tau) = \ell(\tau \circ \sigma^{-1}) + \ell(\sigma)$. This solves Exercise 5.29 **(c)**.

(d) Assume that $\operatorname{Inv}\sigma = \operatorname{Inv}\tau$. We WLOG assume that $\ell(\sigma) \geq \ell(\tau)$ (since otherwise, we can simply switch σ and τ). We have $\operatorname{Inv}\sigma \subseteq \operatorname{Inv}\tau$ (since $\operatorname{Inv}\sigma = \operatorname{Inv}\tau$). But Exercise 5.29 **(c)** shows that $\operatorname{Inv}\sigma \subseteq \operatorname{Inv}\tau$ holds if and only if $\ell(\tau) = \ell(\tau \circ \sigma^{-1}) + \ell(\sigma)$. Hence, we have $\ell(\tau) = \ell(\tau \circ \sigma^{-1}) + \ell(\sigma)$ (since $\operatorname{Inv}\sigma \subseteq \operatorname{Inv}\tau$ holds). Thus, $\ell(\tau) = \ell(\tau \circ \sigma^{-1}) + \underbrace{\ell(\sigma)}_{\geq \ell(\tau)} \geq \ell(\tau \circ \sigma^{-1}) + \ell(\tau)$. Subtracting $\ell(\tau)$ from both sides of this inequality, we obtain $0 \geq \ell(\tau \circ \sigma^{-1})$. In other words, $\ell(\tau \circ \sigma^{-1}) \leq 0$.

But $\ell(\tau \circ \sigma^{-1})$ is the number of inversions of $\tau \circ \sigma^{-1}$ (by the definition of $\ell(\tau \circ \sigma^{-1})$), and thus is a nonnegative integer. Hence, $\ell(\tau \circ \sigma^{-1}) \geq 0$. Combining this with $\ell(\tau \circ \sigma^{-1}) \leq 0$, we obtain $\ell(\tau \circ \sigma^{-1}) = 0$. Thus, Corollary 7.72 (applied to $\tau \circ \sigma^{-1}$ instead of σ) yields $\tau \circ \sigma^{-1} = \operatorname{id}$. Hence, $\underbrace{\tau \circ \sigma^{-1}}_{=\operatorname{id}} \circ \sigma = \operatorname{id} \circ \sigma = \sigma$, so that $\sigma = \tau \circ \underbrace{\sigma^{-1} \circ \sigma}_{=\operatorname{id}} = \tau \circ \operatorname{id} = \tau$. This solves Exercise 5.29 **(d)**.

(e) Let us first prove the logical implication

$$(\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma \circ \tau)) \implies \left((\operatorname{Inv}\sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing \right). \tag{994}$$

[*Proof of (994):* Assume that $\operatorname{Inv}\tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$ holds. We will prove that $(\operatorname{Inv}\sigma) \cap (\operatorname{Inv}(\tau^{-1})) = \varnothing$.

Indeed, fix $c \in (\operatorname{Inv}\sigma) \cap (\operatorname{Inv}(\tau^{-1}))$. Thus, $c \in (\operatorname{Inv}\sigma) \cap (\operatorname{Inv}(\tau^{-1})) \subseteq \operatorname{Inv}\sigma$ and $c \in (\operatorname{Inv}\sigma) \cap (\operatorname{Inv}(\tau^{-1})) \subseteq \operatorname{Inv}(\tau^{-1})$.

We have $c \in \operatorname{Inv}\sigma$. In other words, c is an inversion of σ (since $\operatorname{Inv}\sigma$ is the set of all inversions of σ). In other words, c is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$ (by the definition of an "inversion of σ"). In other words, there exists a pair (i,j) of integers satisfying $1 \leq i < j \leq n$, $\sigma(i) > \sigma(j)$ and $c = (i,j)$. Let us denote this pair (i,j) by (u,v). Thus, (u,v) is a pair of integers satisfying $1 \leq u < v \leq n$, $\sigma(u) > \sigma(v)$ and $c = (u,v)$. We have $v > u$ (since $u < v$). Thus, $\tau(\tau^{-1}(v)) = v > u = \tau(\tau^{-1}(u))$.

We have $(u,v) = c \in \operatorname{Inv}(\tau^{-1})$. In other words, (u,v) is an inversion of τ^{-1} (since $\operatorname{Inv}(\tau^{-1})$ is the set of all inversions of τ^{-1}). In other words, (u,v) is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\tau^{-1}(i) > \tau^{-1}(j)$ (by the definition of an "inversion of τ^{-1}"). In other words, (u,v) is a pair of integers satisfying $1 \leq u < v \leq n$ and $\tau^{-1}(u) > \tau^{-1}(v)$. From $\tau^{-1}(u) > \tau^{-1}(v)$, we obtain $\tau^{-1}(v) < \tau^{-1}(u)$.

We have $\tau^{-1}(u) \in \{1,2,\ldots,n\}$ (since $\tau \in S_n$), so that $\tau^{-1}(u) \leq n$. Also, $\tau^{-1}(v) \in \{1,2,\ldots,n\}$ (since $\tau \in S_n$) and thus $1 \leq \tau^{-1}(v)$. Altogether, we thus know that $(\tau^{-1}(v), \tau^{-1}(u))$ is a pair of integers satisfying $1 \leq \tau^{-1}(v) < \tau^{-1}(u) \leq n$ and $\tau(\tau^{-1}(v)) > \tau(\tau^{-1}(u))$. In other words, $(\tau^{-1}(v), \tau^{-1}(u))$ is a pair (i,j) of integers satisfying $1 \leq i < j \leq n$ and $\tau(i) > \tau(j)$. In other words, $(\tau^{-1}(v), \tau^{-1}(u))$ is an inversion of τ (by the definition of an "inversion of τ"). In other words, $(\tau^{-1}(v), \tau^{-1}(u)) \in \operatorname{Inv}\tau$ (since $\operatorname{Inv}\tau$ is the set of all inversions of τ).

Hence, $\left(\tau^{-1}(v), \tau^{-1}(u)\right) \in \operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$. In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is an inversion of $\sigma \circ \tau$ (since $\operatorname{Inv}(\sigma \circ \tau)$ is the set of all inversions of $\sigma \circ \tau$). In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $(\sigma \circ \tau)(i) > (\sigma \circ \tau)(j)$ (by the definition of an "inversion of $\sigma \circ \tau$"). In other words, $\left(\tau^{-1}(v), \tau^{-1}(u)\right)$ is a pair of integers satisfying $1 \leq \tau^{-1}(v) < \tau^{-1}(u) \leq n$ and $(\sigma \circ \tau)\left(\tau^{-1}(v)\right) > (\sigma \circ \tau)\left(\tau^{-1}(u)\right)$.

In particular, we have $(\sigma \circ \tau)\left(\tau^{-1}(v)\right) > (\sigma \circ \tau)\left(\tau^{-1}(u)\right)$. This rewrites as $\sigma(v) > \sigma(u)$ (since $(\sigma \circ \tau)\left(\tau^{-1}(v)\right) = \left(\sigma \circ \underbrace{\tau \circ \tau^{-1}}_{=\mathrm{id}}\right)(v) = \sigma(v)$ and $(\sigma \circ \tau)\left(\tau^{-1}(u)\right) = \left(\sigma \circ \underbrace{\tau \circ \tau^{-1}}_{=\mathrm{id}}\right)(u) = \sigma(u)$). But this contradicts $\sigma(u) > \sigma(v)$.

Now, forget that we fixed c. We thus have found a contradiction for each $c \in (\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right)$. Hence, there exists no $c \in (\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right)$. In other words, $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right)$ is the empty set. In other words, $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$.

Now, forget our assumption that $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$. We thus have proven that if $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$, then $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$. In other words, we have proven the implication (994).]

Next, we shall prove the logical implication

$$\left((\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing\right) \Longrightarrow (\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)). \tag{995}$$

[*Proof of (995):* Assume that $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$ holds. We will prove that $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$.

Indeed, let $c \in (\operatorname{Inv} \tau) \setminus (\operatorname{Inv}(\sigma \circ \tau))$. We shall prove a contradiction.

We have $c \in (\operatorname{Inv} \tau) \setminus (\operatorname{Inv}(\sigma \circ \tau)) \subseteq \operatorname{Inv} \tau$. In other words, c is an inversion of τ (since $\operatorname{Inv} \tau$ is the set of all inversions of τ). In other words, c is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\tau(i) > \tau(j)$ (by the definition of an "inversion of τ"). In other words, there exists a pair (i, j) of integers satisfying $1 \leq i < j \leq n$, $\tau(i) > \tau(j)$ and $c = (i, j)$. Let us denote this pair (i, j) by (u, v). Thus, (u, v) is a pair of integers satisfying $1 \leq u < v \leq n$, $\tau(u) > \tau(v)$ and $c = (u, v)$. We have $v > u$ (since $u < v$) and $\tau(v) < \tau(u)$ (since $\tau(u) > \tau(v)$).

From $\tau \in S_n$, we obtain $\tau(v) \in \{1, 2, \ldots, n\}$ and thus $1 \leq \tau(v)$. From $\tau \in S_n$, we obtain $\tau(u) \in \{1, 2, \ldots, n\}$ and thus $\tau(u) \leq n$. Also, $\tau^{-1}(\tau(v)) = v > u = \tau^{-1}(\tau(u))$. Altogether, we thus know that $(\tau(v), \tau(u))$ is a pair of integers satisfying $1 \leq \tau(v) < \tau(u) \leq n$ and $\tau^{-1}(\tau(v)) > \tau^{-1}(\tau(u))$. In other words, $(\tau(v), \tau(u))$ is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\tau^{-1}(i) > \tau^{-1}(j)$. In other words, $(\tau(v), \tau(u))$ is an inversion of τ^{-1} (by the definition of an "inversion of τ^{-1}"). In other words, $(\tau(v), \tau(u)) \in \operatorname{Inv}\left(\tau^{-1}\right)$ (since $\operatorname{Inv}\left(\tau^{-1}\right)$ is the set of all inversions of τ^{-1}).

On the other hand, $(u, v) = c \in (\operatorname{Inv} \tau) \setminus (\operatorname{Inv}(\sigma \circ \tau))$, so that $(u, v) \notin \operatorname{Inv}(\sigma \circ \tau)$.

From this, it is easy to obtain that $(\sigma \circ \tau)(u) \leq (\sigma \circ \tau)(v)$ [447]. But the map $\sigma \circ \tau$ is injective[448]. But $u < v$ and therefore $u \neq v$. Hence, $(\sigma \circ \tau)(u) \neq (\sigma \circ \tau)(v)$ (since the map $\sigma \circ \tau$ is injective). Combining this with $(\sigma \circ \tau)(u) \leq (\sigma \circ \tau)(v)$, we obtain $(\sigma \circ \tau)(u) < (\sigma \circ \tau)(v)$. Hence, $\sigma(\tau(u)) = (\sigma \circ \tau)(u) < (\sigma \circ \tau)(v) = \sigma(\tau(v))$. In other words, $\sigma(\tau(v)) > \sigma(\tau(u))$.

Now, we know that $(\tau(v), \tau(u))$ is a pair of integers satisfying $1 \leq \tau(v) < \tau(u) \leq n$ and $\sigma(\tau(v)) > \sigma(\tau(u))$. In other words, $(\tau(v), \tau(u))$ is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $\sigma(i) > \sigma(j)$. In other words, $(\tau(v), \tau(u))$ is an inversion of σ (by the definition of an "inversion of σ"). In other words, $(\tau(v), \tau(u)) \in \operatorname{Inv} \sigma$ (since $\operatorname{Inv} \sigma$ is the set of all inversions of σ). Combining this with $(\tau(v), \tau(u)) \in \operatorname{Inv}\left(\tau^{-1}\right)$, we obtain $(\tau(v), \tau(u)) \in (\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$. Thus, the set $\varnothing$ has at least one element (namely, the element $(\tau(v), \tau(u))$). This contradicts the fact that the set $\varnothing$ has no elements.

Now, forget that we fixed c. We thus have derived a contradiction for each $c \in (\operatorname{Inv} \tau) \setminus (\operatorname{Inv}(\sigma \circ \tau))$. Hence, there exists no $c \in (\operatorname{Inv} \tau) \setminus (\operatorname{Inv}(\sigma \circ \tau))$. In other words, $(\operatorname{Inv} \tau) \setminus (\operatorname{Inv}(\sigma \circ \tau))$ is the empty set. In other words, $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$.

Now, forget our assumption that $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$. We thus have proven that if $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$, then $\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)$. In other words, we have proven the implication (995).]

Combining the two implications (994) and (995), we obtain the logical equivalence

$$(\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)) \iff \left((\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing\right).$$

On the other hand, Exercise 5.29 **(a)** yields that we have the logical equivalence

$$(\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)) \iff (\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)).$$

Hence, we have the following chain of logical equivalences:

$$\begin{aligned}(\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)) &\iff (\operatorname{Inv} \tau \subseteq \operatorname{Inv}(\sigma \circ \tau)) \\ &\iff \left((\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing\right).\end{aligned}$$

In other words, $\ell(\sigma \circ \tau) = \ell(\sigma) + \ell(\tau)$ holds if and only if $(\operatorname{Inv} \sigma) \cap \left(\operatorname{Inv}\left(\tau^{-1}\right)\right) = \varnothing$. This solves Exercise 5.29 **(e)**. □

[447] *Proof.* Assume the contrary. Thus, we don't have $(\sigma \circ \tau)(u) \leq (\sigma \circ \tau)(v)$. Hence, we have $(\sigma \circ \tau)(u) > (\sigma \circ \tau)(v)$. Thus, (u, v) is a pair of integers satisfying $1 \leq u < v \leq n$ and $(\sigma \circ \tau)(u) > (\sigma \circ \tau)(v)$. In other words, (u, v) is a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $(\sigma \circ \tau)(i) > (\sigma \circ \tau)(j)$. In other words, (u, v) is an inversion of $\sigma \circ \tau$ (by the definition of an "inversion of $\sigma \circ \tau$"). In other words, $(u, v) \in \operatorname{Inv}(\sigma \circ \tau)$ (since $\operatorname{Inv}(\sigma \circ \tau)$ is the set of all inversions of $\sigma \circ \tau$). This contradicts $(u, v) \notin \operatorname{Inv}(\sigma \circ \tau)$. This contradiction proves that our assumption was wrong, qed.

[448] *Proof.* We have $\sigma \circ \tau \in S_n$. Hence, $\sigma \circ \tau$ is a permutation of $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\}$). Thus, $\sigma \circ \tau$ is a bijective map. So the map $\sigma \circ \tau$ is bijective, and therefore injective. Qed.

7.68. Solution to Exercise 6.1

Solution to Exercise 6.1. **(a)** We shall solve Exercise 6.1 **(a)** by induction over n:

Induction base: Exercise 6.1 **(a)** holds in the case when $n = 0$ [449]. This completes the induction base.

Induction step: Let k be a positive integer. Assume that Exercise 6.1 **(a)** holds in the case when $n = k - 1$. We must show that Exercise 6.1 **(a)** holds in the case when $n = k$.

Let $\mathbb{K}$ be a commutative ring. Let $a_1, a_2, \ldots, a_k$ be k elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_k$ be k further elements of $\mathbb{K}$. We have assumed that Exercise 6.1 **(a)** holds in the case when $n = k - 1$. Hence, we can apply Exercise 6.1 **(a)** to $k - 1$, $(a_1, a_2, \ldots, a_{k-1})$ and $(b_1, b_2, \ldots, b_{k-1})$ instead of n, $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$. We thus obtain

$$\prod_{i=1}^{k-1} (a_i + b_i) = \sum_{I \subseteq [k-1]} \left(\prod_{i \in I} a_i \right) \left(\prod_{i \in [k-1] \setminus I} b_i \right). \tag{996}$$

Now, k is a positive integer. Thus, we have $k \in [k]$ and $[k] = [k-1] \cup \{k\}$ and $[k-1] = [k] \setminus \{k\}$.

Let us introduce a notation: For any set S, we let $\mathcal{P}(S)$ denote the powerset of S (that is, the set of all subsets of S). Now, we observe the following fact:

Fact 1: Let S be any set. Let $s \in S$. Then, the map

$$\mathcal{P}(S \setminus \{s\}) \to \mathcal{P}(S) \setminus \mathcal{P}(S \setminus \{s\}), \quad U \mapsto U \cup \{s\} \tag{997}$$

is well-defined and a bijection.

[449] *Proof.* Assume that $n = 0$. We must show that Exercise 6.1 **(a)** holds in this case.

We have $n = 0$ and thus $[n] = \varnothing$. Hence, there is only one subset of $[n]$ (namely, $\varnothing$). Therefore,

$$\sum_{I \subseteq [n]} \left(\prod_{i \in I} a_i \right) \left(\prod_{i \in [n] \setminus I} b_i \right) = \underbrace{\left(\prod_{i \in \varnothing} a_i \right)}_{=(\text{empty product})=1} \underbrace{\left(\prod_{i \in [n] \setminus \varnothing} b_i \right)}_{\substack{= \prod_{i \in \varnothing} b_i \\ (\text{since } [n] \setminus \varnothing = [n] = \varnothing)}} = \prod_{i \in \varnothing} b_i = (\text{empty product}) = 1.$$

Comparing this with

$$\prod_{i=1}^{n} (a_i + b_i) = \prod_{i=1}^{0} (a_i + b_i) \qquad (\text{since } n = 0)$$
$$= (\text{empty product}) = 1,$$

we obtain $\prod_{i=1}^{n} (a_i + b_i) = \sum_{I \subseteq [n]} \left(\prod_{i \in I} a_i \right) \left(\prod_{i \in [n] \setminus I} b_i \right)$. Thus, Exercise 6.1 **(a)** holds in the case when $n = 0$.

Fact 1 is easy to prove[450]. We can apply Fact 1 to $S = [k]$ and $s = k$; we thus conclude that the map

$$\mathcal{P}([k] \setminus \{k\}) \to \mathcal{P}([k]) \setminus \mathcal{P}([k] \setminus \{k\}),$$
$$U \mapsto U \cup \{k\}$$

is well-defined and a bijection. Since $[k] \setminus \{k\} = [k-1]$, this rewrites as follows: The map

$$\mathcal{P}([k-1]) \to \mathcal{P}([k]) \setminus \mathcal{P}([k-1]),$$
$$U \mapsto U \cup \{k\}$$

is well-defined and a bijection.

But $[k-1] \subseteq [k]$. Hence, every subset of $[k-1]$ is a subset of $[k]$.

Let us make another helpful observation:

Fact 2: Let I be a subset of $[k-1]$. Then,

$$\prod_{i \in [k] \setminus I} b_i = b_k \prod_{i \in [k-1] \setminus I} b_i \qquad (998)$$

and

$$\prod_{i \in I \cup \{k\}} a_i = a_k \prod_{i \in I} a_i \qquad (999)$$

and

$$\prod_{i \in [k] \setminus (I \cup \{k\})} b_i = \prod_{i \in [k-1] \setminus I} b_i. \qquad (1000)$$

[*Proof of Fact 2:* We have

$$([k] \setminus I) \setminus \{k\} = [k] \setminus \underbrace{(I \cup \{k\})}_{=\{k\} \cup I} = [k] \setminus (\{k\} \cup I) = \underbrace{([k] \setminus \{k\})}_{=[k-1]} \setminus I = [k-1] \setminus I.$$

If we had $k \in I$, then we would have $k \in I \subseteq [k-1]$, which would contradict the fact that $k \notin [k-1]$. Thus, we cannot have $k \in I$. In other words, we have $k \notin I$.

[450] The idea behind it is that the subsets of S which are **not** subsets of $S \setminus \{s\}$ are the subsets of S that contain s, and each such subset can be written in the form $U \cup \{s\}$ for some $U \subseteq S \setminus \{s\}$.

If you want to prove Fact 1 formally, you need to prove two statements:

1. The map (997) is well-defined (i.e., we have $U \cup \{s\} \in \mathcal{P}(S) \setminus \mathcal{P}(S \setminus \{s\})$ for each $U \in \mathcal{P}(S)$).

2. This map is bijective.

Proving the first statement is straightforward. The best way to prove the second statement is to show that the map (997) has an inverse – namely, the map $\mathcal{P}(S) \setminus \mathcal{P}(S \setminus \{s\}) \to \mathcal{P}(S \setminus \{s\})$, $V \mapsto V \setminus \{s\}$. Of course, you would also have to show that this latter map is well-defined, too.

Combining $k \in [k]$ with $k \notin I$, we obtain $k \in [k] \setminus I$. Hence, we can split off the factor for $i = k$ from the product $\prod_{i \in [k] \setminus I} b_i$. We thus obtain

$$\prod_{i \in [k] \setminus I} b_i = b_k \prod_{i \in ([k] \setminus I) \setminus \{k\}} b_i = b_k \prod_{i \in [k-1] \setminus I} b_i$$

(since $([k] \setminus I) \setminus \{k\} = [k-1] \setminus I$). This proves (998).

We have $k \notin I$ and thus $(I \cup \{k\}) \setminus \{k\} = I$.

We have $k \in \{k\} \subseteq I \cup \{k\}$. Thus, we can split off the factor for $i = k$ from the product $\prod_{i \in I \cup \{k\}} a_i$. Thus, we obtain

$$\prod_{i \in I \cup \{k\}} a_i = a_k \prod_{i \in (I \cup \{k\}) \setminus \{k\}} a_i = a_k \prod_{i \in I} a_i$$

(since $(I \cup \{k\}) \setminus \{k\} = I$). This proves (999).

Recall that $[k] \setminus (I \cup \{k\}) = [k-1] \setminus I$. Hence, $\prod_{i \in [k] \setminus (I \cup \{k\})} b_i = \prod_{i \in [k-1] \setminus I} b_i$. This proves (1000). Thus, the proof of Fact 2 is complete.]

We have

$$\begin{aligned}
&\underbrace{\sum_{\substack{I \subseteq [k]; \\ I \in \mathcal{P}([k-1])}}}_{\substack{= \sum_{\substack{I \subseteq [k]; \\ I \subseteq [k-1]}} = \sum_{I \subseteq [k-1]} \\ \text{(since every subset of } [k-1] \\ \text{is a subset of } [k])}} \left(\prod_{i \in I} a_i \right) \left(\prod_{i \in [k] \setminus I} b_i \right) \\
&= \sum_{I \subseteq [k-1]} \left(\prod_{i \in I} a_i \right) \underbrace{\left(\prod_{i \in [k] \setminus I} b_i \right)}_{\substack{= b_k \prod_{i \in [k-1] \setminus I} b_i \\ \text{(by (998))}}} = \sum_{I \subseteq [k-1]} \left(\prod_{i \in I} a_i \right) b_k \left(\prod_{i \in [k-1] \setminus I} b_i \right) \\
&= b_k \underbrace{\sum_{I \subseteq [k-1]} \left(\prod_{i \in I} a_i \right) \left(\prod_{i \in [k-1] \setminus I} b_i \right)}_{\substack{= \prod_{i=1}^{k-1} (a_i + b_i) \\ \text{(by (996))}}} = b_k \prod_{i=1}^{k-1} (a_i + b_i)
\end{aligned} \tag{1001}$$

and

$$\begin{aligned}
&\underbrace{\sum_{\substack{I\subseteq[k];\\ I\notin\mathcal{P}([k-1])}}}_{\substack{=\sum_{\substack{I\in\mathcal{P}([k]);\\ I\notin\mathcal{P}([k-1])}}=\sum_{I\in\mathcal{P}([k])\setminus\mathcal{P}([k-1])}}} \left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k]\setminus I}b_i\right)\\
&=\sum_{I\in\mathcal{P}([k])\setminus\mathcal{P}([k-1])}\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k]\setminus I}b_i\right)\\
&=\sum_{U\in\mathcal{P}([k-1])}\left(\prod_{i\in U\cup\{k\}}a_i\right)\left(\prod_{i\in[k]\setminus(U\cup\{k\})}b_i\right)\\
&\qquad\left(\begin{array}{c}\text{here, we have substituted } U\cup\{k\} \text{ for } I \text{ in the sum,}\\ \text{since the map } \mathcal{P}([k-1])\to\mathcal{P}([k])\setminus\mathcal{P}([k-1]),\ U\mapsto U\cup\{k\}\\ \text{is a bijection}\end{array}\right)\\
&=\sum_{I\in\mathcal{P}([k-1])}\underbrace{\left(\prod_{i\in I\cup\{k\}}a_i\right)}_{\substack{=a_k\prod_{i\in I}a_i\\ \text{(by (999))}}}\underbrace{\left(\prod_{i\in[k]\setminus(I\cup\{k\})}b_i\right)}_{\substack{=\prod_{i\in[k-1]\setminus I}b_i\\ \text{(by (1000))}}}\\
&\qquad(\text{here, we have renamed the summation index } U \text{ as } I)\\
&=\sum_{I\subseteq[k-1]}a_k\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k-1]\setminus I}b_i\right)\\
&=a_k\underbrace{\sum_{I\subseteq[k-1]}\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k-1]\setminus I}b_i\right)}_{\substack{=\prod_{i=1}^{k-1}(a_i+b_i)\\ \text{(by (996))}}}=a_k\prod_{i=1}^{k-1}(a_i+b_i).
\end{aligned}\tag{1002}$$

Now, every subset I of $[k]$ satisfies either $I\in\mathcal{P}([k-1])$ or $I\notin\mathcal{P}([k-1])$ (but not both). Hence,

$$\begin{aligned}
&\sum_{I\subseteq[k]}\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k]\setminus I}b_i\right)\\
&=\underbrace{\sum_{\substack{I\subseteq[k];\\ I\in\mathcal{P}([k-1])}}\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k]\setminus I}b_i\right)}_{\substack{=b_k\prod_{i=1}^{k-1}(a_i+b_i)\\ \text{(by (1001))}}}+\underbrace{\sum_{\substack{I\subseteq[k];\\ I\notin\mathcal{P}([k-1])}}\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k-1]\setminus I}b_i\right)}_{\substack{=a_k\prod_{i=1}^{k-1}(a_i+b_i)\\ \text{(by (1002))}}}\\
&=b_k\prod_{i=1}^{k-1}(a_i+b_i)+a_k\prod_{i=1}^{k-1}(a_i+b_i)=(b_k+a_k)\prod_{i=1}^{k-1}(a_i+b_i)\\
&=(a_k+b_k)\prod_{i=1}^{k-1}(a_i+b_i)=\prod_{i=1}^{k}(a_i+b_i).
\end{aligned}$$

In other words,

$$\prod_{i=1}^{k}(a_i+b_i)=\sum_{I\subseteq[k]}\left(\prod_{i\in I}a_i\right)\left(\prod_{i\in[k]\setminus I}b_i\right). \tag{1003}$$

Now, forget that we fixed $\mathbb{K}$, $(a_1,a_2,\ldots,a_k)$ and $(b_1,b_2,\ldots,b_k)$. We thus have proven (1003) for every commutative ring $\mathbb{K}$, every k elements $a_1,a_2,\ldots,a_k$ of $\mathbb{K}$, and every k elements $b_1,b_2,\ldots,b_k$ of $\mathbb{K}$. In other words, we have proven that Exercise 6.1 **(a)** holds in the case when $n=k$. This completes the induction step. Exercise 6.1 **(a)** is thus proven by induction.

(b) Let $a\in\mathbb{K}$, $b\in\mathbb{K}$ and $n\in\mathbb{N}$. We must prove (338).

For every $k\in\{0,1,\ldots,n\}$, we have

$$(\text{the number of all } I\subseteq[n] \text{ satisfying } |I|=k)=\binom{n}{k} \tag{1004}$$

[451].

[451] *Proof of (1004):* Let $k\in\{0,1,\ldots,n\}$.

Clearly, $[n]$ is an n-element set. Thus, Proposition 3.12 (applied to n, k and $[n]$ instead of m, n and S) shows that $\binom{n}{k}$ is the number of all k-element subsets of $[n]$. In other words,

$$\begin{aligned}
\binom{n}{k}&=(\text{the number of all } k\text{-element subsets of } [n])\\
&=(\text{the number of all } I\subseteq[n] \text{ satisfying } |I|=k).
\end{aligned}$$

This proves (1004).

But Exercise 6.1 **(a)** (applied to $a_i = a$ and $b_i = b$) yields

$$\prod_{i=1}^{n}(a+b) = \sum_{I\subseteq[n]} \underbrace{\left(\prod_{i\in I} a\right)}_{=a^{|I|}} \underbrace{\left(\prod_{i\in[n]\setminus I} b\right)}_{\substack{=b^{|[n]\setminus I|} \\ =b^{|[n]|-|I|} \\ \text{(since } |[n]\setminus I|=|[n]|-|I| \\ \text{(since } I\subseteq[n]))}} = \sum_{I\in\mathcal{P}([n])} a^{|I|}b^{|[n]|-|I|}.$$

Comparing this with $\prod_{i=1}^{n}(a+b) = (a+b)^n$, we obtain

$$\begin{aligned}(a+b)^n &= \underbrace{\sum_{I\subseteq[n]}}_{\substack{=\sum_{k\in\{0,1,\ldots,n\}}\sum_{I\subseteq[n];\ |I|=k} \\ \text{(since every subset } I \text{ of } [n] \\ \text{satisfies } |I|\in\{0,1,\ldots,n\} \\ \text{(because } |I|\leq|[n]|=n))}} a^{|I|}b^{|[n]|-|I|} = \sum_{k\in\{0,1,\ldots,n\}}\sum_{\substack{I\subseteq[n];\\ |I|=k}} \underbrace{a^{|I|}}_{\substack{=a^k \\ \text{(since } |I|=k)}} \underbrace{b^{|[n]|-|I|}}_{\substack{=b^{n-k} \\ \text{(since } |[n]|=n \\ \text{and } |I|=k)}} \\ &= \underbrace{\sum_{k\in\{0,1,\ldots,n\}}}_{=\sum_{k=0}^{n}} \underbrace{\sum_{\substack{I\subseteq[n];\\ |I|=k}} a^k b^{n-k}}_{=(\text{the number of all } I\subseteq[n] \text{ satisfying } |I|=k)a^kb^{n-k}} \\ &= \sum_{k=0}^{n} \underbrace{(\text{the number of all } I\subseteq[n] \text{ satisfying } |I|=k)}_{\substack{=\binom{n}{k} \\ \text{(by (1004))}}} a^kb^{n-k} = \sum_{k=0}^{n}\binom{n}{k}a^kb^{n-k}.\end{aligned}$$

Thus, (338) is proven. This solves Exercise 6.1 **(b)**. □

7.69. Solution to Exercise 6.2

Our solution to Exercise 6.2 will be somewhat similar to our solution to Exercise 3.15 above; in particular, it will rely on Corollary 7.19 again.

In this section, we shall use the notation $\mathbf{m}(k_1, k_2, \ldots, k_m)$ as defined in Exercise 6.2.

Before we solve Exercise 6.2, let us record some really trivial facts:

Lemma 7.155. Each 0-tuple $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$ satisfies $\mathbf{m}(k_1, k_2, \ldots, k_0) = 1$.

(Note that there exists only one 0-tuple $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$, namely the empty list $()$. Thus, the word "each" in Lemma 7.155 is somewhat misleading.)

Proof of Lemma 7.155. Let $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$ be a 0-tuple. Then, the definition of $\mathbf{m}(k_1, k_2, \ldots, k_0)$ yields

$$\begin{aligned}\mathbf{m}(k_1, k_2, \ldots, k_0) &= \frac{(k_1 + k_2 + \cdots + k_0)!}{k_1! k_2! \cdots k_0!} \\ &= \Bigg(\underbrace{k_1 + k_2 + \cdots + k_0}_{=(\text{empty sum})=0}\Bigg)! / \Bigg(\underbrace{k_1! k_2! \cdots k_0!}_{=(\text{empty product})=1}\Bigg) \\ &= 0!/1 = 0! = 1.\end{aligned}$$

This proves Lemma 7.155. □

Lemma 7.156. Let M be a positive integer. Let $(s_1, s_2, \ldots, s_M) \in \mathbb{N}^M$ and $n \in \mathbb{N}$ be such that $s_1 + s_2 + \cdots + s_M = n$. Then,

$$\binom{n}{s_M} \mathbf{m}(s_1, s_2, \ldots, s_{M-1}) = \mathbf{m}(s_1, s_2, \ldots, s_M).$$

Proof of Lemma 7.156. We have

$$(s_1 + s_2 + \cdots + s_{M-1}) + s_M = s_1 + s_2 + \cdots + s_M = n.$$

Subtracting s_M from both sides of this equality, we obtain $s_1 + s_2 + \cdots + s_{M-1} = n - s_M$.

The definition of $\mathbf{m}(s_1, s_2, \ldots, s_{M-1})$ yields

$$\mathbf{m}(s_1, s_2, \ldots, s_{M-1}) = \frac{(s_1 + s_2 + \cdots + s_{M-1})!}{s_1! s_2! \cdots s_{M-1}!} = \frac{(n - s_M)!}{s_1! s_2! \cdots s_{M-1}!} \qquad (1005)$$

(since $s_1 + s_2 + \cdots + s_{M-1} = n - s_M$).

The definition of $\mathbf{m}(s_1, s_2, \ldots, s_M)$ yields

$$\mathbf{m}(s_1, s_2, \ldots, s_M) = \frac{(s_1 + s_2 + \cdots + s_M)!}{s_1! s_2! \cdots s_M!} = \frac{n!}{s_1! s_2! \cdots s_M!} \qquad (1006)$$

(since $s_1 + s_2 + \cdots + s_M = n$).

On the other hand, $n = \underbrace{(s_1 + s_2 + \cdots + s_{M-1})}_{\geq 0} + s_M \geq s_M$. Hence, Proposition 3.4 (applied to n and s_M instead of m and n) yields $\binom{n}{s_M} = \frac{n!}{s_M! (n - s_M)!}$. Multiplying this equality with (1005), we obtain

$$\begin{aligned}&\binom{n}{s_M} \mathbf{m}(s_1, s_2, \ldots, s_{M-1}) \\ &= \frac{n!}{s_M! (n - s_M)!} \cdot \frac{(n - s_M)!}{s_1! s_2! \cdots s_{M-1}!} = \frac{n!}{(s_1! s_2! \cdots s_{M-1}!) s_M!} = \frac{n!}{s_1! s_2! \cdots s_M!} \\ &= \mathbf{m}(s_1, s_2, \ldots, s_M) \qquad (\text{by } (1006)).\end{aligned}$$

This proves Lemma 7.156. □

Solution to Exercise 6.2. We shall solve Exercise 6.2 by induction over m:

Induction base: Exercise 6.2 holds for $m = 0$ [452]. This completes the induction

---

[452] *Proof.* Assume that $m = 0$. We must then show that Exercise 6.2 holds.

From $m = 0$, we obtain $a_1 + a_2 + \cdots + a_m = a_1 + a_2 + \cdots + a_0 = (\text{empty sum}) = 0$.

We are in one of the following two cases:

Case 1: We have $n = 0$.

Case 2: We have $n \neq 0$.

Let us first consider Case 1. In this case, we have $n = 0$. Hence, $0^n = 0^0 = 1$. There exists exactly one 0-tuple $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$ (namely, the empty list $()$), and this 0-tuple $(k_1, k_2, \ldots, k_0)$ satisfies $k_1 + k_2 + \cdots + k_0 = (\text{empty sum}) = 0 = n$. Hence, there exists exactly one 0-tuple $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$ satisfying $k_1 + k_2 + \cdots + k_0 = n$. The sum $\sum\limits_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} 1$ therefore has exactly 1 addend; thus, this sum rewrites as $\sum\limits_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} 1 = 1$.

But recall that $m = 0$. Hence,

$$\sum_{\substack{(k_1,k_2,\ldots,k_m)\in\mathbb{N}^m;\\ k_1+k_2+\cdots+k_m=n}} \mathbf{m}(k_1, k_2, \ldots, k_m) \prod_{i=1}^{m} a_i^{k_i} = \sum_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} \underbrace{\mathbf{m}(k_1, k_2, \ldots, k_0)}_{\substack{=1\\ \text{(by Lemma 7.155)}}} \underbrace{\prod_{i=1}^{0} a_i^{k_i}}_{=(\text{empty product})=1} = \sum_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} 1 = 1.$$

Comparing this with

$$\Bigg(\underbrace{a_1 + a_2 + \cdots + a_m}_{=(\text{empty sum})=0}\Bigg)^n = 0^n = 1,$$

we obtain

$$(a_1 + a_2 + \cdots + a_m)^n = \sum_{\substack{(k_1,k_2,\ldots,k_m)\in\mathbb{N}^m;\\ k_1+k_2+\cdots+k_m=n}} \mathbf{m}(k_1, k_2, \ldots, k_m) \prod_{i=1}^{m} a_i^{k_i}.$$

Hence, Exercise 6.2 holds. We thus have solved Exercise 6.2 in Case 1.

Let us now consider Case 2. In this case, we have $n \neq 0$. Hence, $n > 0$ (since $n \in \mathbb{N}$). Thus, $0^n = 0$.

Each 0-tuple $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$ satisfies $k_1 + k_2 + \cdots + k_0 = (\text{empty sum}) = 0 \neq n$. In other words, no $(k_1, k_2, \ldots, k_0) \in \mathbb{N}^0$ satisfies $k_1 + k_2 + \cdots + k_0 = n$. Hence, the sum $\sum\limits_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} \mathbf{m}(k_1, k_2, \ldots, k_0) \prod\limits_{i=1}^{0} a_i^{k_i}$ is an empty sum. Therefore, this sum rewrites as follows:

$$\sum_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} \mathbf{m}(k_1, k_2, \ldots, k_0) \prod_{i=1}^{0} a_i^{k_i} = (\text{empty sum}) = 0.$$

Now, recall that $m = 0$. Hence,

$$\sum_{\substack{(k_1,k_2,\ldots,k_m)\in\mathbb{N}^m;\\ k_1+k_2+\cdots+k_m=n}} \mathbf{m}(k_1, k_2, \ldots, k_m) \prod_{i=1}^{m} a_i^{k_i} = \sum_{\substack{(k_1,k_2,\ldots,k_0)\in\mathbb{N}^0;\\ k_1+k_2+\cdots+k_0=n}} \mathbf{m}(k_1, k_2, \ldots, k_0) \prod_{i=1}^{0} a_i^{k_i} = 0.$$

base.

Induction step: Fix a positive integer M. Assume that Exercise 6.2 holds for $m = M-1$. We now must show that Exercise 6.2 holds for $m = M$.

We have assumed that Exercise 6.2 holds for $m = M-1$. In other words, the following fact holds:

> *Fact 1:* Let $\mathbb{K}$ be a commutative ring. Let $a_1, a_2, \ldots, a_{M-1}$ be $M-1$ elements of $\mathbb{K}$. Let $n \in \mathbb{N}$. Then,
>
> $$(a_1 + a_2 + \cdots + a_{M-1})^n = \sum_{\substack{(k_1,k_2,\ldots,k_{M-1}) \in \mathbb{N}^{M-1}; \\ k_1+k_2+\cdots+k_{M-1}=n}} \mathbf{m}(k_1, k_2, \ldots, k_{M-1}) \prod_{i=1}^{M-1} a_i^{k_i}.$$

We must show that Exercise 6.2 holds for $m = M$. In other words, we must prove the following fact:

> *Fact 2:* Let $\mathbb{K}$ be a commutative ring. Let $a_1, a_2, \ldots, a_M$ be M elements of $\mathbb{K}$. Let $n \in \mathbb{N}$. Then,
>
> $$(a_1 + a_2 + \cdots + a_M)^n = \sum_{\substack{(k_1,k_2,\ldots,k_M) \in \mathbb{N}^M; \\ k_1+k_2+\cdots+k_M=n}} \mathbf{m}(k_1, k_2, \ldots, k_M) \prod_{i=1}^{M} a_i^{k_i}.$$

[*Proof of Fact 2:* We have

$$a_1 + a_2 + \cdots + a_M = (a_1 + a_2 + \cdots + a_{M-1}) + a_M = a_M + (a_1 + a_2 + \cdots + a_{M-1}).$$

Comparing this with

$$\left(\underbrace{a_1 + a_2 + \cdots + a_m}_{=(\text{empty sum})=0}\right)^n = 0^n = 0,$$

we obtain

$$(a_1 + a_2 + \cdots + a_m)^n = \sum_{\substack{(k_1,k_2,\ldots,k_m) \in \mathbb{N}^m; \\ k_1+k_2+\cdots+k_m=n}} \mathbf{m}(k_1, k_2, \ldots, k_m) \prod_{i=1}^{m} a_i^{k_i}.$$

Hence, Exercise 6.2 holds. We thus have solved Exercise 6.2 in Case 2.

We thus have solved Exercise 6.2 in each of the two Cases 1 and 2. Thus, Exercise 6.2 always holds (under the assumption that $m = 0$). Qed.

Taking both sides of this equality to the n-th power, we obtain

$$\begin{aligned}
&\left(a_1+a_2+\cdots+a_M\right)^n \\
&=\left(a_M+\left(a_1+a_2+\cdots+a_{M-1}\right)\right)^n \\
&=\sum_{k=0}^{n}\binom{n}{k}a_M^k\left(a_1+a_2+\cdots+a_{M-1}\right)^{n-k} \\
&\qquad\left(\text{by (338) (applied to } a=a_M \text{ and } b=a_1+a_2+\cdots+a_{M-1}\text{)}\right) \\
&=\sum_{k=0}^{n}\binom{n}{k}\left(a_1+a_2+\cdots+a_{M-1}\right)^{n-k}a_M^k \\
&=\underbrace{\sum_{r=0}^{n}}_{=\sum_{\substack{r\in\mathbb{N};\\ r\leq n}}}\binom{n}{r}\underbrace{\left(a_1+a_2+\cdots+a_{M-1}\right)^{n-r}}_{\substack{=\sum_{\substack{(k_1,k_2,\ldots,k_{M-1})\in\mathbb{N}^{M-1};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}\mathbf{m}(k_1,k_2,\ldots,k_{M-1})\prod_{i=1}^{M-1}a_i^{k_i}\\ \text{(by Fact 1, applied to } n-r \text{ instead of } r\text{)}}}a_M^r \\
&\qquad\left(\text{here, we have renamed the summation index } k \text{ as } r\right) \\
&=\sum_{\substack{r\in\mathbb{N};\\ r\leq n}}\binom{n}{r}\left(\sum_{\substack{(k_1,k_2,\ldots,k_{M-1})\in\mathbb{N}^{M-1};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}\mathbf{m}\left(k_1,k_2,\ldots,k_{M-1}\right)\prod_{i=1}^{M-1}a_i^{k_i}\right)a_M^r \\
&=\sum_{\substack{r\in\mathbb{N};\\ r\leq n}}\ \sum_{\substack{(k_1,k_2,\ldots,k_{M-1})\in\mathbb{N}^{M-1};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}\binom{n}{r}\mathbf{m}\left(k_1,k_2,\ldots,k_{M-1}\right)\left(\prod_{i=1}^{M-1}a_i^{k_i}\right)a_M^r.
\end{aligned} \tag{1007}$$

On the other hand, each $\left(\left(k_1,k_2,\ldots,k_{M-1}\right),r\right)\in\mathbb{N}^{M-1}\times\mathbb{N}$ satisfying $k_1+k_2+\cdots+k_{M-1}=n-r$ must automatically satisfy $r\leq n$ [453]. Hence, we have the following equality of summation signs:

$$\sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r;\\ r\leq n}}=\sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}.$$

Now, we have the following equality of summation signs:

$$\begin{aligned}
\sum_{\substack{r\in\mathbb{N};\\ r\leq n}}\ \sum_{\substack{(k_1,k_2,\ldots,k_{M-1})\in\mathbb{N}^{M-1};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}&=\sum_{(k_1,k_2,\ldots,k_{M-1})\in\mathbb{N}^{M-1}}\ \sum_{\substack{r\in\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r;\\ r\leq n}} \\
&=\sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r;\\ r\leq n}}=\sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}.
\end{aligned}$$

[453] *Proof.* Let $\left(\left(k_1,k_2,\ldots,k_{M-1}\right),r\right)\in\mathbb{N}^{M-1}\times\mathbb{N}$ be such that $k_1+k_2+\cdots+k_{M-1}=n-r$. Then, $n-r=k_1+k_2+\cdots+k_{M-1}\geq 0$, so that $r\leq n$.

Hence, (1007) becomes

$$\begin{aligned}
&\left(a_1+a_2+\cdots+a_M\right)^n \\
&= \underbrace{\sum_{\substack{r\in\mathbb{N};\\ r\leq n}} \sum_{\substack{(k_1,k_2,\ldots,k_{M-1})\in\mathbb{N}^{M-1};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}}_{=\sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r}}} \dbinom{n}{r} \mathbf{m}\left(k_1,k_2,\ldots,k_{M-1}\right)\left(\prod_{i=1}^{M-1} a_i^{k_i}\right) a_M^r \\
&= \sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r}} \dbinom{n}{r} \mathbf{m}\left(k_1,k_2,\ldots,k_{M-1}\right)\left(\prod_{i=1}^{M-1} a_i^{k_i}\right) a_M^r . \qquad (1008)
\end{aligned}$$

But Corollary 7.19 (applied to $Z=\mathbb{N}$) shows that the map

$$\begin{aligned}
\mathbb{N}^M &\to \mathbb{N}^{M-1}\times\mathbb{N}, \\
\left(s_1,s_2,\ldots,s_M\right) &\mapsto \left(\left(s_1,s_2,\ldots,s_{M-1}\right),s_M\right)
\end{aligned}$$

is a bijection. Hence, we can substitute $\left(\left(s_1,s_2,\ldots,s_{M-1}\right),s_M\right)$ for $\left(\left(k_1,k_2,\ldots,k_{M-1}\right),r\right)$ in the sum on the right hand side of (1008). Thus, we obtain

$$\begin{aligned}
&\sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r}} \dbinom{n}{r} \mathbf{m}\left(k_1,k_2,\ldots,k_{M-1}\right)\left(\prod_{i=1}^{M-1} a_i^{k_i}\right) a_M^r \\
&= \underbrace{\sum_{\substack{(s_1,s_2,\ldots,s_M)\in\mathbb{N}^M;\\ s_1+s_2+\cdots+s_{M-1}=n-s_M}}}_{\substack{=\sum_{\substack{(s_1,s_2,\ldots,s_M)\in\mathbb{N}^M;\\ s_1+s_2+\cdots+s_M=n}}\\ \text{(because for any } (s_1,s_2,\ldots,s_M)\in\mathbb{N}^M\text{,}\\ \text{the condition } (s_1+s_2+\cdots+s_{M-1}=n-s_M)\\ \text{is equivalent to}\\ \text{the condition } (s_1+s_2+\cdots+s_M=n))}} \dbinom{n}{s_M} \mathbf{m}\left(s_1,s_2,\ldots,s_{M-1}\right)\underbrace{\left(\prod_{i=1}^{M-1} a_i^{s_i}\right) a_M^{s_M}}_{=\prod_{i=1}^{M} a_i^{s_i}} \\
&= \sum_{\substack{(s_1,s_2,\ldots,s_M)\in\mathbb{N}^M;\\ s_1+s_2+\cdots+s_M=n}} \underbrace{\dbinom{n}{s_M} \mathbf{m}\left(s_1,s_2,\ldots,s_{M-1}\right)}_{\substack{=\mathbf{m}(s_1,s_2,\ldots,s_M)\\ \text{(by Lemma 7.156)}}} \prod_{i=1}^{M} a_i^{s_i} \\
&= \sum_{\substack{(s_1,s_2,\ldots,s_M)\in\mathbb{N}^M;\\ s_1+s_2+\cdots+s_M=n}} \mathbf{m}\left(s_1,s_2,\ldots,s_M\right)\prod_{i=1}^{M} a_i^{s_i} = \sum_{\substack{(k_1,k_2,\ldots,k_M)\in\mathbb{N}^M;\\ k_1+k_2+\cdots+k_M=n}} \mathbf{m}\left(k_1,k_2,\ldots,k_M\right)\prod_{i=1}^{M} a_i^{k_i}
\end{aligned}$$

(here, we have renamed the summation index $\left(s_1,s_2,\ldots,s_M\right)$ as $\left(k_1,k_2,\ldots,k_M\right)$).

Hence, (1008) becomes

$$\begin{aligned}
&\left(a_1 + a_2 + \cdots + a_M\right)^n \\
&= \sum_{\substack{((k_1,k_2,\ldots,k_{M-1}),r)\in\mathbb{N}^{M-1}\times\mathbb{N};\\ k_1+k_2+\cdots+k_{M-1}=n-r}} \binom{n}{r} \mathbf{m}\left(k_1,k_2,\ldots,k_{M-1}\right) \left(\prod_{i=1}^{M-1} a_i^{k_i}\right) a_M^r \\
&= \sum_{\substack{(k_1,k_2,\ldots,k_M)\in\mathbb{N}^M;\\ k_1+k_2+\cdots+k_M=n}} \mathbf{m}\left(k_1,k_2,\ldots,k_M\right) \prod_{i=1}^{M} a_i^{k_i}.
\end{aligned}$$

This proves Fact 2.]

But Fact 2 is precisely the statement of Exercise 6.2 for $m = M$. Hence, Exercise 6.2 holds for $m = M$ (since Fact 2 is proven). This completes the induction step. Thus, Exercise 6.2 is proven by induction. □

7.70. Solution to Exercise 6.3

Solution to Exercise 6.3. We first notice a purely combinatorial fact: For every $\sigma \in S_n$ satisfying $\sigma \neq \mathrm{id}$,

$$\text{there exists an } i \in \{1,2,\ldots,n\} \text{ such that } \sigma(i) > i \tag{1009}$$

[454]. Thus, for every $\sigma \in S_n$ satisfying $\sigma \neq \mathrm{id}$, we have

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = 0 \tag{1012}$$

[454] *Proof of (1009):* Let $\sigma \in S_n$ be such that $\sigma \neq \mathrm{id}$. We need to prove (1009).

Assume the contrary. Thus, there exists no $i \in \{1,2,\ldots,n\}$ such that $\sigma(i) > i$. In other words, every $i \in \{1,2,\ldots,n\}$ satisfies $\sigma(i) \leq i$.

We shall now show that every $p \in \{1,2,\ldots,n\}$ satisfies

$$\sigma(p) = p. \tag{1010}$$

Proof of (1010): We will prove (1010) by strong induction over p. Thus, fix some $P \in \{1,2,\ldots,n\}$. We assume that (1010) is proven for every $p < P$. We need to show that (1010) holds for $p = P$.

We have assumed that (1010) is proven for every $p < P$. In other words,

$$\sigma(p) = p \qquad \text{for every } p \in \{1,2,\ldots,n\} \text{ satisfying } p < P. \tag{1011}$$

Now, we assume (for the sake of contradiction) that $\sigma(P) \neq P$. Recall that every $i \in \{1,2,\ldots,n\}$ satisfies $\sigma(i) \leq i$. Applying this to $i = P$, we obtain $\sigma(P) \leq P$. Combined with $\sigma(P) \neq P$, this yields $\sigma(P) < P$. Hence, (1011) (applied to $p = \sigma(P)$) yields $\sigma(\sigma(P)) = \sigma(P)$.

But the map σ is a permutation (since $\sigma \in S_n$), thus injective. Hence, from $\sigma(\sigma(P)) = \sigma(P)$, we obtain $\sigma(P) = P$. This contradicts $\sigma(P) \neq P$. Hence, we have obtained a contradiction; thus, our assumption (that $\sigma(P) \neq P$) must have been wrong. We thus have $\sigma(P) = P$. In other words, (1010) holds for $p = P$. This completes the inductive proof of (1010).

Now, (1010) shows that every $p \in \{1,2,\ldots,n\}$ satisfies $\sigma(p) = p = \mathrm{id}(p)$. In other words, $\sigma = \mathrm{id}$. This contradicts $\sigma \neq \mathrm{id}$. This contradiction proves that our assumption was wrong. Hence, (1009) is proven.

[455].

Now, (341) yields

$$\det A = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}$$
$$= \underbrace{(-1)^{\operatorname{id}}}_{=1} \prod_{i=1}^{n} \underbrace{a_{i,\operatorname{id}(i)}}_{=a_{i,i}} + \sum_{\substack{\sigma \in S_n; \\ \sigma \neq \operatorname{id}}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{=0 \\ \text{(by (1012))}}}$$
$$\text{(here, we have moved the addend for } \sigma = \operatorname{id} \text{ out of the sum)}$$
$$= \prod_{i=1}^{n} a_{i,i} + \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma \neq \operatorname{id}}} (-1)^{\sigma} 0}_{=0} = \prod_{i=1}^{n} a_{i,i} = a_{1,1} a_{2,2} \cdots a_{n,n}.$$

This solves Exercise 6.3. □

7.71. Solution to Exercise 6.4

Solution to Exercise 6.4. The map $S_n \to S_n$, $\sigma \mapsto \sigma^{-1}$ (that is, the map from S_n to S_n which sends every permutation to its inverse) is a bijection[456].

Write A in the form $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus, $A^T = \left(a_{j,i}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of A^T). Hence, (341) (applied to A^T and $a_{j,i}$ instead of A and $a_{i,j}$) yields

$$\det\left(A^T\right) = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{\sigma(i),i} = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{\sigma^{-1}(i),i} \qquad (1013)$$

(here, we have substituted σ^{-1} for σ in the sum, since the map $S_n \to S_n$, $\sigma \mapsto \sigma^{-1}$ is a bijection). But every $\sigma \in S_n$ satisfies

$$\prod_{i=1}^{n} a_{\sigma^{-1}(i),i} = \prod_{i=1}^{n} a_{i,\sigma(i)}$$

[455] *Proof of (1012):* Let $\sigma \in S_n$ be such that $\sigma \neq \operatorname{id}$. According to (1009), we know that there exists an $i \in \{1, 2, \ldots, n\}$ such that $\sigma(i) > i$. Let k be such an i. Thus, k is an element of $\{1, 2, \ldots, n\}$ satisfying $\sigma(k) > k$. Hence, $k < \sigma(k)$.

Recall that $a_{i,j} = 0$ for every $(i,j) \in \{1, 2, \ldots, n\}^2$ satisfying $i < j$. Applying this to $i = k$ and $j = \sigma(k)$, we obtain $a_{k,\sigma(k)} = 0$ (since $k < \sigma(k)$). Hence, one factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ is 0 (namely, the factor $a_{k,\sigma(k)}$). Therefore, the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ is 0 (because if one factor of a product is 0, then the whole product is 0).

[456] In fact, this map is its own inverse: Indeed, every $\sigma \in S_n$ satisfies $\left(\sigma^{-1}\right)^{-1} = \sigma$.

[457]. Hence, (1013) becomes

$$\det\left(A^T\right) = \sum_{\sigma\in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^n a_{\sigma^{-1}(i),i}}_{=\prod_{i=1}^n a_{i,\sigma(i)}} = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} = \det A \qquad \text{(by (341))}.$$

This solves Exercise 6.4. □

7.72. Solution to Exercise 6.5

Solution to Exercise 6.5. **(a)** Let u, v and w be three nonnegative integers. Let P be a $u \times v$-matrix, and let Q be a $v \times w$-matrix. We must prove that $(PQ)^T = Q^T P^T$.

Write the $u \times v$-matrix P in the form $P = \left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v}$. Write the $v \times w$-matrix Q in the form $Q = \left(q_{i,j}\right)_{1\leq i\leq v,\ 1\leq j\leq w}$. By the definition of the product PQ, we obtain

$$PQ = \left(\sum_{k=1}^v \underbrace{p_{i,k}q_{k,j}}_{=q_{k,j}p_{i,k}}\right)_{1\leq i\leq u,\ 1\leq j\leq w} \qquad \left(\begin{array}{c}\text{since } P = \left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v} \\ \text{and } Q = \left(q_{i,j}\right)_{1\leq i\leq v,\ 1\leq j\leq w}\end{array}\right)$$
$$= \left(\sum_{k=1}^v q_{k,j}p_{i,k}\right)_{1\leq i\leq u,\ 1\leq j\leq w}.$$

Hence,

$$(PQ)^T = \left(\left(\sum_{k=1}^v q_{k,j}p_{i,k}\right)_{1\leq i\leq u,\ 1\leq j\leq w}\right)^T = \left(\sum_{k=1}^v q_{k,i}p_{j,k}\right)_{1\leq i\leq w,\ 1\leq j\leq u} \qquad (1014)$$

(by the definition of the transpose of a matrix).

On the other hand, from $P = \left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v}$, we obtain

$$P^T = \left(\left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v}\right)^T = \left(p_{j,i}\right)_{1\leq i\leq v,\ 1\leq j\leq u}$$

(by the definition of the transpose of a matrix).

[457]*Proof.* Let $\sigma \in S_n$. Then, σ is a permutation of $\{1, 2, \ldots, n\}$, thus a bijection from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$. Hence, we can substitute $\sigma(i)$ for i in the product $\prod_{i=1}^n a_{\sigma^{-1}(i),i}$. We thus obtain

$$\prod_{i=1}^n a_{\sigma^{-1}(i),i} = \prod_{i=1}^n \underbrace{a_{\sigma^{-1}(\sigma(i)),\sigma(i)}}_{=a_{i,\sigma(i)}} = \prod_{i=1}^n a_{i,\sigma(i)},$$

qed.

Likewise, $Q^T = (q_{j,i})_{1\leq i\leq w,\ 1\leq j\leq v}$.

By the definition of the product $Q^T P^T$, we obtain

$$Q^T P^T = \left(\sum_{k=1}^{v} q_{k,i} p_{j,k}\right)_{1\leq i\leq w,\ 1\leq j\leq u}$$

(since $Q^T = (q_{j,i})_{1\leq i\leq w,\ 1\leq j\leq v}$ and $P^T = (p_{j,i})_{1\leq i\leq v,\ 1\leq j\leq u}$). Comparing this with (1014), we obtain $(PQ)^T = Q^T P^T$. This solves Exercise 6.5 **(a)**.

(b) Let $u \in \mathbb{N}$. We must prove that $(I_u)^T = I_u$.

For any two objects i and j, we define an element $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

Then, it is clear that $\delta_{i,j} = \delta_{j,i}$ for any two objects i and j (since $i = j$ holds if and only if $j = i$). Thus, $(\delta_{i,j})_{1\leq i\leq u,\ 1\leq j\leq u} = (\delta_{j,i})_{1\leq i\leq u,\ 1\leq j\leq u}$.

But $I_u = (\delta_{i,j})_{1\leq i\leq u,\ 1\leq j\leq u}$ (by the definition of I_u). Hence,

$$(I_u)^T = \left((\delta_{i,j})_{1\leq i\leq u,\ 1\leq j\leq u}\right)^T = (\delta_{j,i})_{1\leq i\leq u,\ 1\leq j\leq u}$$

(by the definition of the transpose of a matrix). Comparing this with

$$I_u = (\delta_{i,j})_{1\leq i\leq u,\ 1\leq j\leq u} = (\delta_{j,i})_{1\leq i\leq u,\ 1\leq j\leq u},$$

we obtain $(I_u)^T = I_u$. This solves Exercise 6.5 **(b)**.

(c) Let u and v be two nonnegative integers. Let P be a $u \times v$-matrix. Let $\lambda \in \mathbb{K}$. We must prove that $(\lambda P)^T = \lambda P^T$.

Write the $u \times v$-matrix P in the form $P = (p_{i,j})_{1\leq i\leq u,\ 1\leq j\leq v}$. Thus,

$$P^T = \left((p_{i,j})_{1\leq i\leq u,\ 1\leq j\leq v}\right)^T = (p_{j,i})_{1\leq i\leq v,\ 1\leq j\leq u}$$

(by the definition of the transpose of a matrix). Hence,

$$\lambda \underbrace{P^T}_{=(p_{j,i})_{1\leq i\leq v,\ 1\leq j\leq u}} = \lambda (p_{j,i})_{1\leq i\leq v,\ 1\leq j\leq u} = (\lambda p_{j,i})_{1\leq i\leq v,\ 1\leq j\leq u}. \qquad (1015)$$

On the other hand,

$$\lambda \underbrace{P}_{=(p_{i,j})_{1\leq i\leq u,\ 1\leq j\leq v}} = \lambda (p_{i,j})_{1\leq i\leq u,\ 1\leq j\leq v} = (\lambda p_{i,j})_{1\leq i\leq u,\ 1\leq j\leq v}.$$

Therefore,

$$\begin{aligned}(\lambda P)^T &= \left((\lambda p_{i,j})_{1\leq i\leq u,\ 1\leq j\leq v}\right)^T = (\lambda p_{j,i})_{1\leq i\leq v,\ 1\leq j\leq u} \\ &\qquad \text{(by the definition of the transpose of a matrix)} \\ &= \lambda P^T\end{aligned}$$

(by (1015)). This solves Exercise 6.5 **(c)**.

(d) The solution to Exercise 6.5 **(d)** is very similar to that of Exercise 6.5 **(c)**, and we thus omit it.

(e) Let u and v be two nonnegative integers. Let P be a $u \times v$-matrix. We must prove that $\left(P^T\right)^T = P$.

Write the $u \times v$-matrix P in the form $P = \left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v}$. Thus,

$$P^T = \left(\left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v}\right)^T = \left(p_{j,i}\right)_{1\leq i\leq v,\ 1\leq j\leq u}$$

(by the definition of the transpose of a matrix). Hence,

$$\begin{aligned}\left(P^T\right)^T &= \left(\left(p_{j,i}\right)_{1\leq i\leq v,\ 1\leq j\leq u}\right)^T = \left(p_{i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq v} \\ &\qquad \text{(by the definition of the transpose of a matrix)} \\ &= P.\end{aligned}$$

This solves Exercise 6.5 **(e)**. □

7.73. Solution to Exercise 6.6

Solution to Exercise 6.6. Of course, both parts of Exercise 6.6 can be solved directly using (340). This solution, however, is tedious (particularly for part **(b)** of this exercise). Let us show a smarter way.

(a) Let A be the matrix $\begin{pmatrix} a & b & c & d \\ \ell & 0 & 0 & e \\ k & 0 & 0 & f \\ j & i & h & g \end{pmatrix}$. We want to find $\det A$.

We write the matrix A in the form $A = \left(a_{u,v}\right)_{1\leq u\leq 4,\ 1\leq v\leq 4}$ [458]. Thus,

$$\begin{array}{llll} a_{1,1} = a, & a_{1,2} = b, & a_{1,3} = c, & a_{1,4} = d, \\ a_{2,1} = \ell, & a_{2,2} = 0, & a_{2,3} = 0, & a_{2,4} = e, \\ a_{3,1} = k, & a_{3,2} = 0, & a_{3,3} = 0, & a_{3,4} = f, \\ a_{4,1} = j, & a_{4,2} = i, & a_{4,3} = h, & a_{4,4} = g. \end{array}$$

Now, applying (340) to $n = 4$, we obtain

$$\det A = \sum_{\sigma \in S_4} (-1)^\sigma a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,\sigma(3)} a_{4,\sigma(4)}. \qquad (1016)$$

The sum on the right hand side of (1016) has $|S_4| = 4! = 24$ addends. However, some of them are 0. Namely, every addend corresponding to a permutation

[458] We cannot write $A = \left(a_{i,j}\right)_{1\leq i\leq 4,\ 1\leq j\leq 4}$ because the letters i and j are already taken for something different.

$\sigma \in S_4$ satisfying $\sigma(2) \notin \{1,4\}$ must be 0 [459]. Hence, all such addends can be removed from the sum (without changing the value of this sum). Similarly, all addends corresponding to permutations $\sigma \in S_4$ satisfying $\sigma(3) \notin \{1,4\}$ must be 0, and can therefore also be removed from the sum. The addends that survive these two removals are the ones that correspond to permutations $\sigma \in S_4$ satisfying $\sigma(2) \in \{1,4\}$ and $\sigma(3) \in \{1,4\}$. It is easy to see that there are exactly four such permutations: In one-line notation, these permutations are $(2,1,4,3)$, $(2,4,1,3)$, $(3,1,4,2)$ and $(3,4,1,2)$. The addends corresponding to these permutations are $a_{1,2}a_{2,1}a_{3,4}a_{4,3}$, $-a_{1,2}a_{2,4}a_{3,1}a_{4,3}$, $-a_{1,3}a_{2,1}a_{3,4}a_{4,2}$ and $a_{1,3}a_{2,4}a_{3,1}a_{4,2}$. Hence, (1016) simplifies to

$$\begin{aligned} &\det A \\ &= \underbrace{a_{1,2}}_{=b}\underbrace{a_{2,1}}_{=\ell}\underbrace{a_{3,4}}_{=f}\underbrace{a_{4,3}}_{=h} - \underbrace{a_{1,2}}_{=b}\underbrace{a_{2,4}}_{=e}\underbrace{a_{3,1}}_{=k}\underbrace{a_{4,3}}_{=h} - \underbrace{a_{1,3}}_{=c}\underbrace{a_{2,1}}_{=\ell}\underbrace{a_{3,4}}_{=f}\underbrace{a_{4,2}}_{=i} + \underbrace{a_{1,3}}_{=c}\underbrace{a_{2,4}}_{=e}\underbrace{a_{3,1}}_{=k}\underbrace{a_{4,2}}_{=i} \\ &= b\ell fh - bekh - c\ell fi + ceki. \end{aligned}$$

This is a simple enough formula to consider an answer to Exercise 6.6 **(a)**, but we can simplify it even further. Namely,

$$\det A = b\ell fh - bekh - c\ell fi + ceki = (bh - ci)(\ell f - ek).$$

Exercise 6.6 **(a)** is solved.

(b) Let A be the matrix $\begin{pmatrix} a & b & c & d & e \\ f & 0 & 0 & 0 & g \\ h & 0 & 0 & 0 & i \\ j & 0 & 0 & 0 & k \\ \ell & m & n & o & p \end{pmatrix}$. We want to find $\det A$.

We write the matrix A in the form $A = (a_{u,v})_{1\leq u\leq 5,\ 1\leq v\leq 5}$. Thus, $a_{1,1} = a$, $a_{1,2} = b$, etc.. For us, the most important property of A is that the 3×3-submatrix in the middle of A is filled with zeroes. In other words,

$$a_{u,v} = 0 \qquad \text{for every } u \in \{2,3,4\} \text{ and } v \in \{2,3,4\}. \tag{1017}$$

Now, applying (340) to $n = 5$, we obtain

$$\det A = \sum_{\sigma\in S_5} (-1)^{\sigma} a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,\sigma(3)}a_{4,\sigma(4)}a_{5,\sigma(5)}. \tag{1018}$$

[459] *Proof.* Let $\sigma \in S_4$ be such that $\sigma(2) \notin \{1,4\}$. We must then show that the addend on the right hand side of (1016) corresponding to this σ must be 0. In other words, we have to show that $(-1)^{\sigma} a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,\sigma(3)}a_{4,\sigma(4)} = 0$.

We have $\sigma(2) \notin \{1,4\}$, and thus $\sigma(2) \in \{2,3\}$. Hence, $a_{2,\sigma(2)} = 0$ (because $a_{2,2} = 0$ and $a_{2,3} = 0$), and thus $(-1)^{\sigma} a_{1,\sigma(1)} \underbrace{a_{2,\sigma(2)}}_{=0} a_{3,\sigma(3)}a_{4,\sigma(4)} = 0$, qed.

But every $\sigma \in S_5$ satisfies $a_{2,\sigma(2)}a_{3,\sigma(3)}a_{4,\sigma(4)} = 0$ [460]. Hence, (1018) becomes

$$\det A = \sum_{\sigma \in S_5} (-1)^\sigma a_{1,\sigma(1)} \underbrace{a_{2,\sigma(2)}a_{3,\sigma(3)}a_{4,\sigma(4)}}_{=0} a_{5,\sigma(5)} = \sum_{\sigma \in S_5} (-1)^\sigma a_{1,\sigma(1)} 0 a_{5,\sigma(5)} = 0.$$

Exercise 6.6 **(b)** is thus solved. □

7.74. Solution to Exercise 6.7

Our solution to Exercise 6.7 relies on Lemma 6.17. Thus, the reader is advised to read the proof of said lemma before the solution. Furthermore, we shall use the following simple fact:

Lemma 7.157. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Then, the columns of A are the transposes of the respective rows of A^T.

Proof of Lemma 7.157. Fix $k \in \{1, 2, \ldots, m\}$.

Write the $n \times m$-matrix A in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Then, $A^T = (a_{j,i})_{1 \leq i \leq m,\ 1 \leq j \leq n}$ (by the definition of A^T). Hence,

$$\left(\text{the } k\text{-th row of } A^T\right) = (a_{j,k})_{1 \leq i \leq 1,\ 1 \leq j \leq n}.$$

Thus,

$$\begin{aligned} &\left(\text{the transpose of } \underbrace{\text{the } k\text{-th row of } A^T}_{=(a_{j,k})_{1 \leq i \leq 1,\ 1 \leq j \leq n}}\right) \\ &= \left(\text{the transpose of } (a_{j,k})_{1 \leq i \leq 1,\ 1 \leq j \leq n}\right) = \left((a_{j,k})_{1 \leq i \leq 1,\ 1 \leq j \leq n}\right)^T \\ &= (a_{i,k})_{1 \leq i \leq n,\ 1 \leq j \leq 1} \end{aligned} \tag{1019}$$

[460] *Proof.* Let $\sigma \in S_5$. Then, σ is a permutation of $\{1, 2, 3, 4, 5\}$, and thus an injective map. Therefore, the numbers $\sigma(2), \sigma(3), \sigma(4)$ are pairwise distinct.

We now claim that there exists an $u \in \{2, 3, 4\}$ such that $\sigma(u) \in \{2, 3, 4\}$. In order to prove this, we assume the contrary. Thus, every $u \in \{2, 3, 4\}$ satisfies $\sigma(u) \notin \{2, 3, 4\}$. Hence, every $u \in \{2, 3, 4\}$ satisfies $\sigma(u) \in \{1, 5\}$ (since $\sigma(u) \in \{1, 2, 3, 4, 5\}$ but $\sigma(u) \notin \{2, 3, 4\}$). In other words, the numbers $\sigma(2), \sigma(3), \sigma(4)$ belong to $\{1, 5\}$. Hence, $\sigma(2), \sigma(3), \sigma(4)$ are three distinct numbers belonging to the set $\{1, 5\}$. But this is absurd, since the set $\{1, 5\}$ does not contain three distinct numbers. Hence, we have obtained a contradiction. This shows that our assumption was wrong.

We thus have shown that there exists an $u \in \{2, 3, 4\}$ such that $\sigma(u) \in \{2, 3, 4\}$. Consider such a u. Applying (1017) to $v = \sigma(u)$, we now obtain $a_{u,\sigma(u)} = 0$. But $u \in \{2, 3, 4\}$, so that $a_{u,\sigma(u)}$ is a factor in the product $a_{2,\sigma(2)}a_{3,\sigma(3)}a_{4,\sigma(4)}$. Hence, the product $a_{2,\sigma(2)}a_{3,\sigma(3)}a_{4,\sigma(4)}$ is 0 (since its factor $a_{u,\sigma(u)}$ is 0), qed.

(by the definition of $\left(\left(a_{j,k}\right)_{1\leq i\leq 1,\ 1\leq j\leq n}\right)^T$). On the other hand, we have $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$, and thus

$$\begin{aligned}(\text{the } k\text{-th column of } A) &= \left(a_{i,k}\right)_{1\leq i\leq n,\ 1\leq j\leq 1} \\ &= \left(\text{the transpose of the } k\text{-th row of } A^T\right) \qquad (1020)\end{aligned}$$

(by (1019)).

Now, forget that we fixed k. We thus have proven that (1020) holds for each $k \in \{1, 2, \ldots, m\}$. In other words, for each $k \in \{1, 2, \ldots, m\}$, the k-th column of A is the transpose of the k-th row of A^T. In other words, the columns of A are the transposes of the respective rows of A^T. This proves Lemma 7.157. □

Remark 7.158. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Lemma 7.157 (applied to $m = n$) shows that the columns of A are the transposes of the respective rows of A^T. Thus, we have a correspondence between the columns of A and the rows of A^T. We can use this correspondence to "transport" information about the columns of A to the rows of A^T and vice versa; for example:

- If a column of A consists of zeroes, then the corresponding row of A^T consists of zeroes. The converse also holds.
- If two columns of A are equal, then the corresponding two rows of A^T are equal. The converse also holds.
- Let $k \in \{1, 2, \ldots, n\}$ and $\lambda \in \mathbb{K}$. If B is the $n \times n$-matrix obtained from A by multiplying the k-th column by λ, then B^T is the $n \times n$-matrix obtained from A^T by multiplying the k-th row by λ. Again, the converse also holds.
- Let $k \in \{1, 2, \ldots, n\}$. Let A' be an $n \times n$-matrix whose columns equal the corresponding columns of A except (perhaps) the k-th column. Then, $(A')^T$ is an $n \times n$-matrix whose rows equal the corresponding rows of A^T except (perhaps) the k-th row. Again, the converse also holds.
- Let $k \in \{1, 2, \ldots, n\}$. Let A' be a further $n \times n$-matrix. Let B be the $n \times n$-matrix obtained from A by adding the k-th column of A' to the k-th column of A. Then, B^T is the $n \times n$-matrix obtained from A^T by adding the k-th row of $(A')^T$ to the k-th row of A'. Again, the converse holds as well.

Solution to Exercise 6.7. Let us write the matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Thus, every $(i,j) \in \{1, 2, \ldots, n\}^2$ satisfies

$$(\text{the } (i,j)\text{-th entry of the matrix } A) = a_{i,j}. \qquad (1021)$$

We let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

(a) Let B be an $n \times n$-matrix obtained from A by switching two rows. Thus, there exist two distinct elements u and v of $\{1, 2, \ldots, n\}$ such that B is the $n \times n$-matrix obtained from A by switching the u-th row with the v-th row. Consider these u and v.

We write the matrix B in the form $B = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Consider the transposition $t_{u,v}$ in S_n (defined according to Definition 5.29). Clearly, $t_{u,v}$ is a permutation of the set $\{1, 2, \ldots, n\} = [n]$, thus a map $[n] \to [n]$. Also, $(-1)^{t_{u,v}} = -1$ (by Exercise 5.10 **(b)**, applied to $i = u$ and $j = v$).

Recall that B is the $n \times n$-matrix obtained from A by switching the u-th row with the v-th row. In other words, for all $k \in \{1, 2, \ldots, n\}$, we have

$$(\text{the } k\text{-th row of } B) = (\text{the } t_{u,v}(k) \text{-th row of } A) \tag{1022}$$

(because $t_{u,v}$ is the permutation of $\{1, 2, \ldots, n\}$ that switches u with v while leaving all other numbers fixed). Therefore, every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfies

$$b_{i,j} = a_{t_{u,v}(i),j}$$

[461]. Hence, $B = \left(\underbrace{b_{i,j}}_{= a_{t_{u,v}(i),j}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(a_{t_{u,v}(i),j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Therefore, we can apply Lemma 6.17 **(a)** to $t_{u,v}$, A, $a_{i,j}$ and B instead of κ, B, $b_{i,j}$ and B_κ. We thus obtain $\det B = \underbrace{(-1)^{t_{u,v}}}_{=-1} \cdot \det A = -\det A$. Exercise 6.7 **(a)** is thus solved.

(b) Let B be an $n \times n$-matrix obtained from A by switching two columns. Thus, B^T is an $n \times n$-matrix obtained from A^T by switching two rows (because the columns of A correspond to the rows of A^T [462]). Hence, Exercise 6.7 **(a)** (applied to A^T and B^T instead of A and B) yields $\det\left(B^T\right) = -\det\left(A^T\right)$.

---

[461] *Proof.* We have $B = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus, every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfies

$$(\text{the } (i, j) \text{-th entry of the matrix } B) = b_{i,j}. \tag{1023}$$

Now, let $(i, j) \in \{1, 2, \ldots, n\}^2$. Then,

$$\begin{aligned} b_{i,j} &= (\text{the } (i, j) \text{-th entry of the matrix } B) \qquad (\text{by (1023)}) \\ &= \left(\text{the } j\text{-th entry of } \underbrace{\text{the } i\text{-th row of } B}_{\substack{= (\text{the } t_{u,v}(i)\text{-th row of } A) \\ (\text{by (1022), applied to } k = i)}} \right) \\ &= (\text{the } j\text{-th entry of the } t_{u,v}(i) \text{-th row of } A) \\ &= a_{t_{u,v}(i),j} \qquad (\text{by (1021), applied to } t_{u,v}(i) \text{ instead of } i), \end{aligned}$$

qed.

[462] See Remark 7.158 for the meaning of "correspond" we are using here.

But Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Also, Exercise 6.4 (applied to B instead of A) yields $\det\left(B^T\right) = \det B$. But recall that $\det\left(B^T\right) = -\det\left(A^T\right)$. This rewrites as $\det B = -\det A$ (since $\det\left(B^T\right) = \det B$ and $\det\left(A^T\right) = \det A$). This solves Exercise 6.7 **(b)**.

(c) Assume that a row of A consists of zeroes. Thus, there exists a $u \in \{1, 2, \ldots, n\}$ such that the u-th row of A consists of zeroes. Consider this u.

The u-th row of the matrix A consists of zeroes. In other words,

$$a_{u,j} = 0 \qquad \text{for every } j \in \{1, 2, \ldots, n\}. \tag{1024}$$

Now, every $\sigma \in S_n$ satisfies $\prod_{i=1}^{n} a_{i,\sigma(i)} = 0$ [463]. Thus, (341) shows that

$$\det A = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^{n} a_{i,\sigma(i)}}_{=0} = \sum_{\sigma \in S_n} (-1)^\sigma 0 = 0.$$

This solves Exercise 6.7 **(c)**.

(d) Assume that a column of A consists of zeroes. Thus, a row of A^T consists of zeroes (because the columns of A correspond to the rows of A^T [464]). Hence, Exercise 6.7 **(c)** (applied to A^T instead of A) yields $\det\left(A^T\right) = 0$.

But Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Hence, $\det A = \det\left(A^T\right) = 0$. This solves Exercise 6.7 **(d)**.

(e) Assume that A has two equal rows. In other words, some two distinct rows of A are equal (where "distinct" means that these rows are in different positions, not that they are distinct as row vectors). In other words, there exist two distinct elements u and v of $\{1, 2, \ldots, n\}$ such that

$$(\text{the } u\text{-th row of } A) = (\text{the } v\text{-th row of } A). \tag{1025}$$

Consider these u and v.

The equality (1025) shows that

$$a_{u,j} = a_{v,j} \qquad \text{for every } j \in \{1, 2, \ldots, n\} \tag{1026}$$

(because $a_{u,j}$ is the j-th entry of the u-th row of A, while $a_{v,j}$ is the j-th entry of the v-th row of A).

Define a map $\kappa : [n] \to [n]$ by

$$\left(\kappa(i) = \begin{cases} i, & \text{if } i \neq u; \\ v, & \text{if } i = u \end{cases} \qquad \text{for every } i \in [n]\right).$$

[463] *Proof.* Let $\sigma \in S_n$. Then, the u-th factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ is $a_{u,\sigma(u)} = 0$ (by (1024), applied to $j = \sigma(u)$). Hence, the whole product is 0. In other words, we have $\prod_{i=1}^{n} a_{i,\sigma(i)} = 0$, qed.

[464] See Remark 7.158 for the meaning of "correspond" we are using here.

The definition of κ shows that $\kappa(v) = v$ (since $v \neq u$) but also $\kappa(u) = v$. Thus, $\kappa(u) = v = \kappa(v)$, in spite of $u \neq v$. Therefore, the map κ is not injective, and thus not bijective; in particular, κ is not a permutation. Thus, $\kappa \notin S_n$. But every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfy

$$a_{\kappa(i),j} = a_{i,j} \tag{1027}$$

[465]. Thus, $\left(\underbrace{a_{\kappa(i),j}}_{=a_{i,j}}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n} = A$. Therefore, we can apply Lemma 6.17 **(b)** to A, $a_{i,j}$ and A instead of B, $b_{i,j}$ and B_κ. We thus obtain $\det A = 0$. Exercise 6.7 **(e)** is thus solved.

(f) Assume that A has two equal columns. Thus, A^T has two equal rows (because the columns of A correspond to the rows of A^T [466]). Hence, Exercise 6.7 **(e)** (applied to A^T instead of A) yields $\det(A^T) = 0$.

But Exercise 6.4 yields $\det(A^T) = \det A$. Hence, $\det A = \det(A^T) = 0$. This solves Exercise 6.7 **(f)**.

(g) Let B be the $n \times n$-matrix obtained from A by multiplying the k-th row by λ. Thus,

$$(\text{the } k\text{-th row of } B) = \lambda (\text{the } k\text{-th row of } A), \tag{1028}$$

whereas

$$((\text{the } u\text{-th row of } B) = (\text{the } u\text{-th row of } A) \text{ for all } u \in \{1, 2, \ldots, n\} \text{ satisfying } u \neq k). \tag{1029}$$

We write the matrix B in the form $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$. Thus, every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfies

$$(\text{the } (i, j)\text{-th entry of the matrix } B) = b_{i,j}. \tag{1030}$$

[465] *Proof of (1027):* Let $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$. We must prove (1027). We must be in one of the following two cases:

Case 1: We have $i \neq u$.

Case 2: We have $i = u$.

Let us first consider Case 1. In this case, we have $i \neq u$. Thus, the definition of κ yields $\kappa(i) = \begin{cases} i, & \text{if } i \neq u; \\ v, & \text{if } i = u \end{cases} = i$ (since $i \neq u$), so that $a_{\kappa(i),j} = a_{i,j}$. Thus, (1027) is proven in Case 1.

Let us now consider Case 2. In this case, we have $i = u$. Thus, $\kappa(i) = \kappa(u) = v$ (since $i = u$). Hence,

$$\begin{aligned} a_{\kappa(i),j} = a_{v,j} &= a_{u,j} && (\text{by (1026)}) \\ &= a_{i,j} && (\text{since } u = i). \end{aligned}$$

Thus, (1027) is proven in Case 2.

We have now proven (1027) in both Cases 1 and 2. Thus, (1027) always holds, qed.

[466] See Remark 7.158 for the meaning of "correspond" we are using here.

For every $u \in \{1, 2, \ldots, n\}$ and $v \in \{1, 2, \ldots, n\}$ satisfying $u \neq k$, we have

$$b_{u,v} = a_{u,v} \tag{1031}$$

[467]. For every $v \in \{1, 2, \ldots, n\}$, we have

$$b_{k,v} = \lambda a_{k,v} \tag{1032}$$

[468]. Now, it is easy to see that every $\sigma \in S_n$ satisfies

$$\prod_{i=1}^{n} b_{i,\sigma(i)} = \lambda \prod_{i=1}^{n} a_{i,\sigma(i)} \tag{1033}$$

[469]. Now, recall that $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence, (341) (applied to B and $b_{i,j}$

---

[467] *Proof of (1031):* Let $u \in \{1, 2, \ldots, n\}$ and $v \in \{1, 2, \ldots, n\}$ be such that $u \neq k$. Then, (1029) shows that the v-th entry of the u-th row of B equals the v-th entry of the u-th row of A. Since the former entry is $b_{u,v}$, while the latter entry equals $a_{u,v}$, this rewrites as $b_{u,v} = a_{u,v}$. This proves (1031).

[468] *Proof of (1032):* Let $v \in \{1, 2, \ldots, n\}$. Then, (1028) shows that the v-th entry of the k-th row of B equals λ times the v-th entry of the k-th row of A. Since the former entry is $b_{k,v}$, while the latter entry equals $a_{k,v}$, this rewrites as $b_{k,v} = \lambda a_{k,v}$. This proves (1032).

[469] *Proof of (1033):* Let $\sigma \in S_n$. Taking the k-th factor out of the product $\prod\limits_{i=1}^{n} a_{i,\sigma(i)}$, we obtain

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = a_{k,\sigma(k)} \cdot \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} a_{i,\sigma(i)}.$$

Multiplying both sides of this equality by λ, we obtain

$$\lambda \prod_{i=1}^{n} a_{i,\sigma(i)} = \lambda a_{k,\sigma(k)} \cdot \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} a_{i,\sigma(i)}.$$

Compared with

$$\begin{aligned}
\prod_{i=1}^{n} b_{i,\sigma(i)} &= \underbrace{b_{k,\sigma(k)}}_{\substack{=\lambda a_{k,\sigma(k)}\\ \text{(by (1032), applied to}\\ v=\sigma(k))}} \cdot \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} \underbrace{b_{i,\sigma(i)}}_{\substack{=a_{i,\sigma(i)}\\ \text{(by (1031), applied to}\\ u=i \text{ and } v=\sigma(i))}} \qquad (\text{since } k \in \{1, 2, \ldots, n\}) \\
&= \lambda a_{k,\sigma(k)} \cdot \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} a_{i,\sigma(i)},
\end{aligned}$$

this yields $\prod\limits_{i=1}^{n} b_{i,\sigma(i)} = \lambda \prod\limits_{i=1}^{n} a_{i,\sigma(i)}$. This proves (1033).

instead of A and $a_{i,j}$) yields

$$\det B = \sum_{\sigma \in S_n} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} b_{i,\sigma(i)}}_{\substack{=\lambda \prod_{i=1}^{n} a_{i,\sigma(i)} \\ \text{(by (1033))}}} = \sum_{\sigma \in S_n} (-1)^{\sigma} \lambda \prod_{i=1}^{n} a_{i,\sigma(i)}$$
$$= \lambda \underbrace{\sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{=\det A \\ \text{(by (341))}}} = \lambda \det A.$$

Thus, Exercise 6.7 **(g)** is solved.

(h) Let B be the $n \times n$-matrix obtained from A by multiplying the k-th column by λ. Thus, B^T is the $n \times n$-matrix obtained from A^T by multiplying the k-th row by λ (because the columns of A correspond to the rows of A^T [470]). Hence, Exercise 6.7 **(g)** (applied to A^T and B^T instead of A and B) yields $\det\left(B^T\right) = \lambda \det\left(A^T\right)$.

But Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Also, Exercise 6.4 (applied to B instead of A) yields $\det\left(B^T\right) = \det B$. But recall that $\det\left(B^T\right) = \lambda \det\left(A^T\right)$. This rewrites as $\det B = \lambda \det A$ (since $\det\left(B^T\right) = \det B$ and $\det\left(A^T\right) = \det A$). This solves Exercise 6.7 **(h)**.

(i) We know that the rows of the matrix A' equal the corresponding rows of A except (perhaps) the k-th row. In other words,

$$\left(\left(\text{the } u\text{-th row of } A'\right) = (\text{the } u\text{-th row of } A) \quad \text{for all } u \in \{1,2,\ldots,n\} \text{ satisfying } u \neq k\right). \tag{1034}$$

We know that B is the $n \times n$-matrix obtained from A by adding the k-th row of A' to the k-th row of A. Hence,

$$(\text{the } k\text{-th row of } B) = (\text{the } k\text{-th row of } A) + \left(\text{the } k\text{-th row of } A'\right) \tag{1035}$$

and

$$((\text{the } u\text{-th row of } B) = (\text{the } u\text{-th row of } A) \quad \text{for all } u \in \{1,2,\ldots,n\} \text{ satisfying } u \neq k). \tag{1036}$$

We write the matrix B in the form $B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, and we write the matrix A' in the form $A' = \left(a'_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Then, for every $u \in \{1,2,\ldots,n\}$ and $v \in \{1,2,\ldots,n\}$ satisfying $u \neq k$, we have

$$a'_{u,v} = a_{u,v} \tag{1037}$$

[470] See Remark 7.158 for the meaning of "correspond" we are using here.

[471] and

$$b_{u,v} = a_{u,v} \tag{1038}$$

[472]. Also, for every $v \in \{1, 2, \ldots, n\}$, we have

$$b_{k,v} = a_{k,v} + a'_{k,v} \tag{1039}$$

[473]. Now, it is easy to see that every $\sigma \in S_n$ satisfies

$$\prod_{i=1}^{n} b_{i,\sigma(i)} = \prod_{i=1}^{n} a_{i,\sigma(i)} + \prod_{i=1}^{n} a'_{i,\sigma(i)} \tag{1040}$$

[474].

We have $A = \left(a'_{i,j}\right)_{1 \le i \le n,\ 1 \le j \le n}$; therefore, (341) (applied to A' and $a'_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det A' = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a'_{i,\sigma(i)}. \tag{1043}$$

Now, recall that $B = (b_{i,j})_{1 \le i \le n,\ 1 \le j \le n}$. Hence, (341) (applied to B and $b_{i,j}$ instead

---

[471] This follows from (1034) in a similar way as (1031) follows from (1029).

[472] This follows from (1036) in a similar way as (1031) follows from (1029).

[473] *Proof of (1039):* Let $v \in \{1, 2, \ldots, n\}$. Then, (1035) shows that

$$\begin{aligned} &(\text{the } v\text{-th entry of the } k\text{-th row of } B) \\ &= (\text{the } v\text{-th entry of the } k\text{-th row of } A) + \left(\text{the } v\text{-th entry of the } k\text{-th row of } A'\right). \end{aligned}$$

But the three entries appearing in this equality are $b_{k,v}$, $a_{k,v}$ and $a'_{k,v}$ (in this order). Thus, this equality rewrites as $b_{k,v} = a_{k,v} + a'_{k,v}$. This proves (1039).

[474] *Proof of (1040):* Let $\sigma \in S_n$. Pulling the k-th factor out of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$, we obtain

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = a_{k,\sigma(k)} \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} a_{i,\sigma(i)} \tag{1041}$$

(since $k \in \{1, 2, \ldots, n\}$). Similarly,

$$\prod_{i=1}^{n} a'_{i,\sigma(i)} = a'_{k,\sigma(k)} \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} \underbrace{a'_{i,\sigma(i)}}_{\substack{= a_{i,\sigma(i)} \\ \text{(by (1037), applied to} \\ u = i \text{ and } v = \sigma(i))}} = a'_{k,\sigma(k)} \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} a_{i,\sigma(i)} \tag{1042}$$

of A and $a_{i,j}$) yields

$$\det B = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^n b_{i,\sigma(i)}}_{\substack{= \prod_{i=1}^n a_{i,\sigma(i)} + \prod_{i=1}^n a'_{i,\sigma(i)} \\ \text{(by (1040))}}} = \sum_{\sigma \in S_n} (-1)^\sigma \left(\prod_{i=1}^n a_{i,\sigma(i)} + \prod_{i=1}^n a'_{i,\sigma(i)} \right)$$
$$= \underbrace{\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}}_{\substack{= \det A \\ \text{(by (341))}}} + \underbrace{\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^n a'_{i,\sigma(i)}}_{\substack{= \det A' \\ \text{(by (1043))}}} = \det A + \det A'.$$

This solves Exercise 6.7 **(i)**.

(j) We know that A' is an $n \times n$-matrix whose columns equal the corresponding columns of A except (perhaps) the k-th column. Thus, $(A')^T$ is an $n \times n$-matrix whose rows equal the corresponding rows of A^T except (perhaps) the k-th row (because the columns of A correspond to the rows of A^T [475], and similarly the columns of A' correspond to the rows of $(A')^T$).

Also, we know that B is the $n \times n$-matrix obtained from A by adding the k-th column of A' to the k-th column of A. Hence, B^T is the $n \times n$-matrix obtained from A^T by adding the k-th row of $(A')^T$ to the k-th row of A^T (because the columns of A correspond to the rows of A^T, and similarly the columns of A' correspond to the rows of $(A')^T$, and similarly the columns of B correspond to the rows of B^T).

Hence, Exercise 6.7 **(i)** (applied to A^T, $(A')^T$ and B^T instead of A, A' and B) yields $\det(B^T) = \det(A^T) + \det((A')^T)$.

---

and

$$\prod_{i=1}^n b_{i,\sigma(i)} = \underbrace{b_{k,\sigma(k)}}_{\substack{= a_{k,\sigma(k)} + a'_{k,\sigma(k)} \\ \text{(by (1039), applied to} \\ v = \sigma(k))}} \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} \underbrace{b_{i,\sigma(i)}}_{\substack{= a_{i,\sigma(i)} \\ \text{(by (1038), applied to} \\ u = i \text{ and } v = \sigma(i))}}$$
$$= \left(a_{k,\sigma(k)} + a'_{k,\sigma(k)} \right) \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} a_{i,\sigma(i)}$$
$$= \underbrace{a_{k,\sigma(k)} \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} a_{i,\sigma(i)}}_{\substack{= \prod_{i=1}^n a_{i,\sigma(i)} \\ \text{(by (1041))}}} + \underbrace{a'_{k,\sigma(k)} \cdot \prod_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq k}} a_{i,\sigma(i)}}_{\substack{= \prod_{i=1}^n a'_{i,\sigma(i)} \\ \text{(by (1042))}}}$$
$$= \prod_{i=1}^n a_{i,\sigma(i)} + \prod_{i=1}^n a'_{i,\sigma(i)}.$$

This proves (1040).

[475]See Remark 7.158 for the meaning of "correspond" we are using here.

But Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Also, Exercise 6.4 (applied to B instead of A) yields $\det\left(B^T\right) = \det B$. Finally, Exercise 6.4 (applied to A' instead of A) yields $\det\left(\left(A'\right)^T\right) = \det A'$.

But recall that $\det\left(B^T\right) = \det\left(A^T\right) + \det\left(\left(A'\right)^T\right)$. This rewrites as $\det B = \det A + \det A'$ (since $\det\left(B^T\right) = \det B$ and $\det\left(A^T\right) = \det A$ and $\det\left(\left(A'\right)^T\right) = \det A'$). This solves Exercise 6.7 **(j)**.

[*Remark:* It is tempting to regard Exercise 6.7 **(e)** as a consequence of Exercise 6.7 **(a)**, because if a matrix A has two equal rows, then switching these two rows does not change the matrix A, and thus Exercise 6.7 **(a)** (applied to $B = A$) yields that $\det A = -\det A$ in this case. However, we cannot conclude $\det A = 0$ from $\det A = -\det A$ in general, unless we know that we can "divide by 2" (or at least cancel a factor of 2 from equalities) in the commutative ring $\mathbb{K}$. For example, if $\mathbb{K}$ is the ring $\mathbb{Z}/2\mathbb{Z}$, then every $a \in \mathbb{K}$ satisfies $a = -a$, but not every $a \in \mathbb{K}$ satisfies $a = 0$. Of course, if $\mathbb{K} = \mathbb{Z}$ or $\mathbb{K} = \mathbb{Q}$ or $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, then this argument works and thus Exercise 6.7 **(e)** does follow from Exercise 6.7 **(a)** in these cases.] □

7.75. Solution to Exercise 6.8

Our solution to Exercise 6.8 relies on Lemma 6.17 and on Lemma 7.157. Thus, we advise the reader to read the proofs of these two lemmas before the solution of the exercise.

Before we start solving Exercise 6.8, let us prove a simple fact:

Lemma 7.159. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix such that some row of the matrix A equals a scalar multiple of some other row of A. Then,

$$\det A = 0. \tag{1044}$$

Proof of Lemma 7.159. Some row of the matrix A equals a scalar multiple of some other row of A. In other words, there exist two distinct elements k and ℓ of $\{1, 2, \ldots, n\}$ such that the k-th row of the matrix A equals a scalar multiple of the ℓ-th row of A. Consider these k and ℓ.

Let C be the $n \times n$-matrix obtained from A by replacing the k-th row of A by the ℓ-th row of A. Then, the matrix C has two equal rows (namely, its k-th row and its ℓ-th row). Hence, $\det C = 0$ (by Exercise 6.7 **(e)**, applied to C instead of A).

Recall that the k-th row of the matrix A equals a scalar multiple of the ℓ-th row of A. In other words, there exists a $\lambda \in \mathbb{K}$ such that

$$(\text{the } k\text{-th row of } A) = \lambda\,(\text{the } \ell\text{-th row of } A)\,. \tag{1045}$$

Consider this λ. By the construction of C, we have

$$(\text{the } k\text{-th row of } C) = (\text{the } \ell\text{-th row of } A)\,. \tag{1046}$$

Thus, (1045) becomes

$$(\text{the } k\text{-th row of } A) = \lambda \underbrace{(\text{the } \ell\text{-th row of } A)}_{\substack{=(\text{the } k\text{-th row of } C) \\ (\text{by (1046)})}} = \lambda (\text{the } k\text{-th row of } C). \tag{1047}$$

On the other hand, for every $u \in \{1, 2, \ldots, n\}$ satisfying $u \neq k$, we have

$$(\text{the } u\text{-th row of } C) = (\text{the } u\text{-th row of } A)$$

(since the construction of C involves modifying only the k-th row of A) and thus

$$(\text{the } u\text{-th row of } A) = (\text{the } u\text{-th row of } C). \tag{1048}$$

Now, combining (1047) with (1048), we obtain the following description of A from C: The matrix A is the matrix obtained from C by multiplying the k-th row by λ. Then, $\det A = \lambda \det C$ (by Exercise 6.7 **(g)**, applied to C and A instead of A and B). Thus, $\det A = \lambda \underbrace{\det C}_{=0} = 0$. This proves Lemma 7.159. □

Now, let us come to the actual solution of Exercise 6.8.

Solution to Exercise 6.8. Let us write the $n \times n$-matrix A in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

(a) We need to prove that if we add a scalar multiple of a row of A to another row of A, then the determinant of A does not change. In other words, we need to prove that if B is an $n \times n$-matrix obtained from A by adding a scalar multiple of a row of A to another row of A, then $\det B = \det A$.

So let B be an $n \times n$-matrix obtained from A by adding a scalar multiple of a row of A to another row of A. We then need to show that $\det B = \det A$.

We know that B is an $n \times n$-matrix obtained from A by adding a scalar multiple of a row of A to another row of A. In other words, there exist two distinct elements ℓ and k of $\{1, 2, \ldots, n\}$ such that B is the $n \times n$-matrix obtained from A by adding a scalar multiple of the ℓ-th row of A to the k-th row of A. Consider these ℓ and k.

We know that B is the $n \times n$-matrix obtained from A by adding a scalar multiple of the ℓ-th row of A to the k-th row of A. In other words, there exists a $\lambda \in \mathbb{K}$ such that B is $n \times n$-matrix obtained from A by adding $\lambda \cdot (\text{the } \ell\text{-th row of } A)$ to the k-th row of A. Consider this λ.

Let A' be the $n \times n$-matrix obtained from A by replacing the k-th row of A by $\lambda \cdot (\text{the } \ell\text{-th row of } A)$. Then,

$$\begin{aligned}(\text{the } k\text{-th row of } A') &= \lambda \cdot \underbrace{(\text{the } \ell\text{-th row of } A)}_{\substack{=(\text{the } \ell\text{-th row of } A') \\ (\text{since the } \ell\text{-th row of } A' \text{ has been} \\ \text{taken over from } A \text{ unchanged})}} \\ &= \lambda \cdot (\text{the } \ell\text{-th row of } A').\end{aligned} \tag{1049}$$

Thus, the k-th row of A' equals a scalar multiple of the ℓ-th row of A'. Thus, some row of the matrix A' equals a scalar multiple of some other row of A'. Hence, (1044) (applied to A' instead of A) shows that $\det A' = 0$.

On the other hand, A' is an $n \times n$-matrix whose rows equal the corresponding rows of A except (perhaps) the k-th row (because of the construction of A'). Also, B is the $n \times n$-matrix obtained from A by adding $\lambda \cdot (\text{the } \ell\text{-th row of } A)$ to the k-th row of A. Because of (1049), this can be rewritten as follows: B is the $n \times n$-matrix obtained from A by adding the k-th row of A' to the k-th row of A. Exercise 6.7 **(i)** thus yields $\det B = \det A + \underbrace{\det A'}_{=0} = \det A$.

Thus, we have shown that

$$\det B = \det A. \tag{1050}$$

This completes our solution to Exercise 6.8 **(a)**.

(b) We need to prove that if we add a scalar multiple of a column of A to another column of A, then the determinant of A does not change. In other words, we need to prove that if B is an $n \times n$-matrix obtained from A by adding a scalar multiple of a column of A to another column of A, then $\det B = \det A$.

So let B be an $n \times n$-matrix obtained from A by adding a scalar multiple of a column of A to another column of A. We then need to show that $\det B = \det A$.

We know that B is an $n \times n$-matrix obtained from A by adding a scalar multiple of a column of A to another column of A. Therefore, B^T is an $n \times n$-matrix obtained from A^T by adding a scalar multiple of a row of A^T to another row of A^T (because the columns of A correspond to the rows of A^T [476]). Therefore, (1050) (applied to A^T and B^T instead of A and B) yields $\det\left(B^T\right) = \det\left(A^T\right)$.

But Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Also, Exercise 6.4 (applied to B instead of A) yields $\det\left(B^T\right) = \det B$. But recall that $\det\left(B^T\right) = \det\left(A^T\right)$. This rewrites as $\det B = \det A$ (since $\det\left(B^T\right) = \det B$ and $\det\left(A^T\right) = \det A$). This solves Exercise 6.8 **(b)**. □

7.76. Solution to Exercise 6.9

Our goal is now to prove Lemma 6.20. Actually, we will prove a more general result:

> **Lemma 7.160.** Let $n \in \mathbb{N}$. For every $i \in \{1, 2, \ldots, n\}$, let Z_i be a finite set. For every $i \in \{1, 2, \ldots, n\}$ and every $k \in Z_i$, let $p_{i,k}$ be an element of $\mathbb{K}$. Then,
>
> $$\prod_{i=1}^{n} \sum_{k \in Z_i} p_{i,k} = \sum_{(k_1, k_2, \ldots, k_n) \in Z_1 \times Z_2 \times \cdots \times Z_n} \prod_{i=1}^{n} p_{i,k_i}.$$

Let us first show the particular case of this lemma for $n = 2$:

[476] See Remark 7.158 for the meaning of "correspond" we are using here.

Lemma 7.161. Let X and Y be two finite sets. For every $x \in X$, let q_x be an element of $\mathbb{K}$. For every $y \in Y$, let r_y be an element of $\mathbb{K}$. Then,

$$\left(\sum_{x\in X} q_x\right)\left(\sum_{y\in Y} r_y\right) = \sum_{(x,y)\in X\times Y} q_x r_y .$$

Proof of Lemma 7.161. We have

$$\underbrace{\sum_{(x,y)\in X\times Y}}_{=\sum_{x\in X}\sum_{y\in Y}} q_x r_y = \sum_{x\in X} \underbrace{\sum_{y\in Y} q_x r_y}_{=q_x \sum_{y\in Y} r_y} = \sum_{x\in X}\left(q_x \sum_{y\in Y} r_y\right) = \left(\sum_{x\in X} q_x\right)\left(\sum_{y\in Y} r_y\right).$$

This proves Lemma 7.161. □

Proof of Lemma 7.160. We claim that

$$\prod_{i=1}^{m} \sum_{k\in Z_i} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_m)\in Z_1\times Z_2\times\cdots\times Z_m} \prod_{i=1}^{m} p_{i,k_i}. \tag{1051}$$

for every $m \in \{0,1,\ldots,n\}$.

[*Proof of (1051):* We shall prove (1051) by induction over m:

Induction base: Comparing

$$\prod_{i=1}^{0} \sum_{k\in Z_i} p_{i,k} = (\text{empty product}) = 1$$

with

$$\begin{aligned}\sum_{(k_1,k_2,\ldots,k_0)\in Z_1\times Z_2\times\cdots\times Z_0} \underbrace{\prod_{i=1}^{0} p_{i,k_i}}_{=(\text{empty product})=1} &= \sum_{(k_1,k_2,\ldots,k_0)\in Z_1\times Z_2\times\cdots\times Z_0} 1 \\ &= \underbrace{\left|Z_1\times Z_2\times\cdots\times Z_0\right|}_{\substack{=1\\ (\text{since } Z_1\times Z_2\times\cdots\times Z_0 \text{ is an}\\ \text{empty Cartesian product})}} \cdot 1 = 1,\end{aligned}$$

we obtain $\prod_{i=1}^{0} \sum_{k\in Z_i} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_0)\in Z_1\times Z_2\times\cdots\times Z_0} \prod_{i=1}^{0} p_{i,k_i}$. In other words, (1051) holds for $m = 0$. This completes the induction base.

Induction step: Let $M \in \{0,1,\ldots,n\}$ be positive. Assume that (1051) holds for $m = M-1$. We now must show that (1051) holds for $m = M$.

We have assumed that (1051) holds for $m = M - 1$. In other words, we have

$$\prod_{i=1}^{M-1} \sum_{k\in Z_i} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_{M-1})\in Z_1\times Z_2\times\cdots\times Z_{M-1}} \prod_{i=1}^{M-1} p_{i,k_i}. \tag{1052}$$

Lemma 7.18 shows that the map

$$Z_1 \times Z_2 \times \cdots \times Z_M \to (Z_1 \times Z_2 \times \cdots \times Z_{M-1}) \times Z_M,$$
$$(s_1, s_2, \ldots, s_M) \mapsto ((s_1, s_2, \ldots, s_{M-1}), s_M)$$

is a bijection.

For every $(k_1, k_2, \ldots, k_{M-1}) \in Z_1 \times Z_2 \times \cdots \times Z_{M-1}$, we define an element $g_{(k_1,k_2,\ldots,k_{M-1})} \in \mathbb{K}$ by

$$g_{(k_1,k_2,\ldots,k_{M-1})} = \prod_{i=1}^{M-1} p_{i,k_i}. \tag{1053}$$

Now, (1052) becomes

$$\begin{aligned}
\prod_{i=1}^{M-1} \sum_{k\in Z_i} p_{i,k} &= \sum_{(k_1,k_2,\ldots,k_{M-1})\in Z_1\times Z_2\times\cdots\times Z_{M-1}} \underbrace{\prod_{i=1}^{M-1} p_{i,k_i}}_{\substack{=g_{(k_1,k_2,\ldots,k_{M-1})} \\ \text{(by (1053))}}} \\
&= \sum_{(k_1,k_2,\ldots,k_{M-1})\in Z_1\times Z_2\times\cdots\times Z_{M-1}} g_{(k_1,k_2,\ldots,k_{M-1})} \\
&= \sum_{x\in Z_1\times Z_2\times\cdots\times Z_{M-1}} g_x
\end{aligned} \tag{1054}$$

(here, we have renamed the summation index $(k_1, k_2, \ldots, k_{M-1})$ as x).

The sets $Z_1, Z_2, \ldots, Z_{M-1}$ are finite. Hence, their Cartesian product $Z_1 \times Z_2 \times \cdots \times Z_{M-1}$ is also finite. Every $(s_1, s_2, \ldots, s_M) \in Z_1 \times Z_2 \times \cdots \times Z_M$ satisfies

$$\prod_{i=1}^{M} p_{i,s_i} = \left(\prod_{i=1}^{M-1} p_{i,s_i}\right) p_{M,s_M} \tag{1055}$$

(indeed, this follows by splitting off the factor for $i = M$ from the product $\prod_{i=1}^{M} p_{i,s_i}$).

Now, if we split off the factor for $i = M$ from the product $\prod_{i=1}^{M} \sum_{k \in Z_i} p_{i,k}$, we obtain

$$\prod_{i=1}^{M} \sum_{k \in Z_i} p_{i,k} = \underbrace{\left(\prod_{i=1}^{M-1} \sum_{k \in Z_i} p_{i,k} \right)}_{\substack{= \sum\limits_{x \in Z_1 \times Z_2 \times \cdots \times Z_{M-1}} g_x \\ \text{(by (1054))}}} \underbrace{\left(\sum_{k \in Z_M} p_{M,k} \right)}_{\substack{= \sum\limits_{y \in Z_M} p_{M,y} \\ \text{(here, we renamed the} \\ \text{summation index } k \text{ as } y)}}$$

$$= \left(\sum_{x \in Z_1 \times Z_2 \times \cdots \times Z_{M-1}} g_x \right) \left(\sum_{y \in Z_M} p_{M,y} \right) = \sum_{(x,y) \in (Z_1 \times Z_2 \times \cdots \times Z_{M-1}) \times Z_M} g_x p_{M,y}$$

$$\left(\begin{array}{c} \text{by Lemma 7.161, applied to} \\ X = Z_1 \times Z_2 \times \cdots \times Z_{M-1},\ Y = Z_M,\ q_x = g_x \text{ and } r_y = p_{M,y} \end{array} \right)$$

$$= \sum_{(s_1,s_2,\ldots,s_M) \in Z_1 \times Z_2 \times \cdots \times Z_M} \underbrace{g_{(s_1,s_2,\ldots,s_{M-1})}}_{\substack{= \prod\limits_{i=1}^{M-1} p_{i,s_i} \\ \text{(by the definition of } g_{(s_1,s_2,\ldots,s_{M-1})})}} p_{M,s_M}$$

$$\left(\begin{array}{c} \text{here, we have substituted } ((s_1,s_2,\ldots,s_{M-1}),s_M) \text{ for } (x,y) \\ \text{in the sum, since the map} \\ Z_1 \times Z_2 \times \cdots \times Z_M \to (Z_1 \times Z_2 \times \cdots \times Z_{M-1}) \times Z_M, \\ (s_1,s_2,\ldots,s_M) \mapsto ((s_1,s_2,\ldots,s_{M-1}),s_M) \\ \text{is a bijection} \end{array} \right)$$

$$= \sum_{(s_1,s_2,\ldots,s_M) \in Z_1 \times Z_2 \times \cdots \times Z_M} \underbrace{\left(\prod_{i=1}^{M-1} p_{i,s_i} \right) p_{M,s_M}}_{\substack{= \prod\limits_{i=1}^{M} p_{i,s_i} \\ \text{(by (1055))}}}$$

$$= \sum_{(s_1,s_2,\ldots,s_M) \in Z_1 \times Z_2 \times \cdots \times Z_M} \prod_{i=1}^{M} p_{i,s_i} = \sum_{(k_1,k_2,\ldots,k_M) \in Z_1 \times Z_2 \times \cdots \times Z_M} \prod_{i=1}^{M} p_{i,k_i}$$

(here, we have renamed the summation index $(s_1, s_2, \ldots, s_M)$ as $(k_1, k_2, \ldots, k_M)$). In other words, (1051) holds for $m = M$. This completes the induction step. Thus, (1051) is proven by induction.]

Now, (1051) (applied to $m = n$) yields

$$\prod_{i=1}^{n} \sum_{k \in Z_i} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_n) \in Z_1 \times Z_2 \times \cdots \times Z_n} \prod_{i=1}^{n} p_{i,k_i}.$$

This proves Lemma 7.160. □

Proof of Lemma 6.20. For every $i \in [n]$, the set $[m_i]$ is clearly a finite set. For every $i \in [n]$, we know that $p_{i,1}, p_{i,2}, \ldots, p_{i,m_i}$ are elements of $\mathbb{K}$. In other words, for every

$i \in [n]$ and every $k \in [m_i]$, we know that $p_{i,k}$ is an element of $\mathbb{K}$. In other words, for every $i \in \{1, 2, \ldots, n\}$ and every $k \in [m_i]$, we know that $p_{i,k}$ is an element of $\mathbb{K}$ (since $[n] = \{1, 2, \ldots, n\}$). Hence, Lemma 7.160 (applied to $Z_i = [m_i]$) yields

$$\prod_{i=1}^{n} \sum_{k \in [m_i]} p_{i,k} = \sum_{(k_1,k_2,\ldots,k_n) \in [m_1] \times [m_2] \times \cdots \times [m_n]} \prod_{i=1}^{n} p_{i,k_i}.$$

Thus,

$$\sum_{(k_1,k_2,\ldots,k_n) \in [m_1] \times [m_2] \times \cdots \times [m_n]} \prod_{i=1}^{n} p_{i,k_i} = \prod_{i=1}^{n} \underbrace{\sum_{k \in [m_i]}}_{= \sum_{k=1}^{m_i}} p_{i,k} = \prod_{i=1}^{n} \sum_{k=1}^{m_i} p_{i,k}.$$

This proves Lemma 6.20. □

We have now proven Lemma 6.20, and thus solved Exercise 6.9.

We note that Lemma 7.160 generalizes the well-known fact that any $n \in \mathbb{N}$ and any n finite sets $A_1, A_2, \ldots, A_n$ satisfy

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_n|.$$

Indeed, if $n \in \mathbb{N}$ and if $A_1, A_2, \ldots, A_n$ are n finite sets, then Lemma 7.160 (applied to $Z_i = A_i$ and $p_{i,k} = 1$) yields

$$\prod_{i=1}^{n} \sum_{k \in A_i} 1 = \sum_{(k_1,k_2,\ldots,k_n) \in A_1 \times A_2 \times \cdots \times A_n} \underbrace{\prod_{i=1}^{n} 1}_{=1} = \sum_{(k_1,k_2,\ldots,k_n) \in A_1 \times A_2 \times \cdots \times A_n} 1$$
$$= |A_1 \times A_2 \times \cdots \times A_n| \cdot 1 = |A_1 \times A_2 \times \cdots \times A_n|$$

and thus

$$|A_1 \times A_2 \times \cdots \times A_n| = \prod_{i=1}^{n} \underbrace{\sum_{k \in A_i} 1}_{= |A_i| \cdot 1 = |A_i|} = \prod_{i=1}^{n} |A_i| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_n|.$$

7.77. Solution to Exercise 6.10

We shall solve Exercise 6.10 more or less as you would expect, by modifying our proof of (369) in some places.

Solution to Exercise 6.10. Let B be the 2×2-matrix $\begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix}$. Thus,

$$\det B = \det \begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix} = a \cdot 0 - 1 \cdot b = -b.$$

Let A be the 2×2-matrix $\begin{pmatrix} x_{k+2} & x_{k+1} \\ x_1 & x_0 \end{pmatrix}$. Then, $\det A = \det \begin{pmatrix} x_{k+2} & x_{k+1} \\ x_1 & x_0 \end{pmatrix} = x_{k+2}x_0 - x_{k+1}x_1$.

We now claim that

$$AB^m = \begin{pmatrix} x_{m+k+2} & x_{m+k+1} \\ x_{m+1} & x_m \end{pmatrix} \qquad \text{for every } m \in \mathbb{N}. \tag{1056}$$

[*Proof of (1056):* We shall prove (1056) by induction over m:

Induction base: We have $A \underbrace{B^0}_{=I_2} = AI_2 = A = \begin{pmatrix} x_{k+2} & x_{k+1} \\ x_1 & x_0 \end{pmatrix}$. Compared with $\begin{pmatrix} x_{0+k+2} & x_{0+k+1} \\ x_{0+1} & x_0 \end{pmatrix} = \begin{pmatrix} x_{k+2} & x_{k+1} \\ x_1 & x_0 \end{pmatrix}$, this yields $AB^0 = \begin{pmatrix} x_{0+k+2} & x_{0+k+1} \\ x_{0+1} & x_0 \end{pmatrix}$. In other words, (1056) holds for $m = 0$. This completes the induction base.

Induction step: Let M be a positive integer. Assume that (1056) holds for $m = M - 1$. We need to show that (1056) holds for $m = M$.

We have assumed that (1056) holds for $m = M - 1$. In other words,

$$AB^{M-1} = \begin{pmatrix} x_{(M-1)+k+2} & x_{(M-1)+k+1} \\ x_{(M-1)+1} & x_{M-1} \end{pmatrix} = \begin{pmatrix} x_{M+k+1} & x_{M+k} \\ x_M & x_{M-1} \end{pmatrix}.$$

Now,

$$\begin{aligned} A \underbrace{B^M}_{=B^{M-1} \cdot B} &= \underbrace{AB^{M-1}}_{=\begin{pmatrix} x_{M+k+1} & x_{M+k} \\ x_M & x_{M-1} \end{pmatrix}} \cdot \underbrace{B}_{=\begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix}} \\ &= \begin{pmatrix} x_{M+k+1} & x_{M+k} \\ x_M & x_{M-1} \end{pmatrix} \cdot \begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix} \\ &= \begin{pmatrix} x_{M+k+1} \cdot a + x_{M+k} \cdot b & x_{M+k+1} \cdot 1 + x_{M+k} \cdot 0 \\ x_M \cdot a + x_{M-1} \cdot b & x_M \cdot 1 + x_{M-1} \cdot 0 \end{pmatrix} \\ &\qquad \text{(by the definition of a product of two matrices)} \\ &= \begin{pmatrix} ax_{M+k+1} + bx_{M+k} & x_{M+k+1} \\ ax_M + bx_{M-1} & x_M \end{pmatrix}. \end{aligned} \tag{1057}$$

But (371) (applied to $n = M + k + 2$) yields

$$x_{M+k+2} = a \underbrace{x_{(M+k+2)-1}}_{=x_{M+k+1}} + b \underbrace{x_{(M+k+2)-2}}_{=x_{M+k}} = ax_{M+k+1} + bx_{M+k}.$$

Also, (371) (applied to $n = M + 1$) yields $x_{M+1} = a \underbrace{x_{(M+1)-1}}_{=x_M} + b \underbrace{x_{(M+1)-2}}_{=x_{M-1}} = ax_M + bx_{M-1}$. Now,

$$\begin{pmatrix} x_{M+k+2} & x_{M+k+1} \\ x_{M+1} & x_M \end{pmatrix} = \begin{pmatrix} ax_{M+k+1} + bx_{M+k} & x_{M+k+1} \\ ax_M + bx_{M-1} & x_M \end{pmatrix}$$

(since $x_{M+k+2} = ax_{M+k+1} + bx_{M+k}$ and $x_{M+1} = ax_M + bx_{M-1}$). Compared with (1057), this yields $AB^M = \begin{pmatrix} x_{M+k+2} & x_{M+k+1} \\ x_{M+1} & x_M \end{pmatrix}$. In other words, (1056) holds for $m = M$. This completes the induction step. Thus, (1056) is proven by induction.]

Now, let $n > k$ be an integer. Then, $n - k > 0$, so that $n - k - 1 \in \mathbb{N}$. Hence, (1056) (applied to $m = n - k - 1$) yields

$$AB^{n-k-1} = \begin{pmatrix} x_{(n-k-1)+k+2} & x_{(n-k-1)+k+1} \\ x_{(n-k-1)+1} & x_{n-k-1} \end{pmatrix} = \begin{pmatrix} x_{n+1} & x_n \\ x_{n-k} & x_{n-k-1} \end{pmatrix}.$$

Taking determinants on both sides of this equality, we obtain

$$\det\left(AB^{n-k-1}\right) = \det\begin{pmatrix} x_{n+1} & x_n \\ x_{n-k} & x_{n-k-1} \end{pmatrix} = x_{n+1}x_{n-k-1} - x_n x_{n-k}.$$

Hence,

$$\begin{aligned} & x_{n+1}x_{n-k-1} - x_n x_{n-k} \\ &= \det\left(AB^{n-k-1}\right) = \det A \cdot \underbrace{\det\left(B^{n-k-1}\right)}_{\substack{=(\det B)^{n-k-1} \\ \text{(by Corollary 6.25 (b), applied} \\ \text{to 2 and } n-k-1 \text{ instead of } n \text{ and } k)}} \\ & \qquad \left(\text{by Theorem 6.23, applied to 2 and } B^{n-k-1} \text{ instead of } n \text{ and } B\right) \\ &= \underbrace{\det A}_{=x_{k+2}x_0 - x_{k+1}x_1} \cdot \Bigg(\underbrace{\det B}_{=-b}\Bigg)^{n-k-1} = \left(x_{k+2}x_0 - x_{k+1}x_1\right) \cdot (-b)^{n-k-1} \\ &= (-b)^{n-k-1}\left(x_{k+2}x_0 - x_{k+1}x_1\right). \end{aligned}$$

This solves Exercise 6.10. □

7.78. Solution to Exercise 6.11

Solution to Exercise 6.11. We let B be the $n \times n$-matrix $\left(\dbinom{i-1}{j-1}\right)_{1\le i\le n,\ 1\le j\le n}$. We have $\dbinom{i-1}{j-1} = 0$ for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$ [477]. Therefore,

[477] *Proof.* Let $(i,j) \in \{1,2,\ldots,n\}^2$ be such that $i < j$. Then, $i - 1 \in \mathbb{N}$ (since $i \geq 1$) and $j - 1 \in \mathbb{N}$ (since $j \geq 1$) and $i - 1 < j - 1$ (since $i < j$). Hence, (231) (applied to $i - 1$ and $j - 1$ instead of m and n) yields $\dbinom{i-1}{j-1} = 0$, qed.

Exercise 6.3 (applied to B and $\binom{i-1}{j-1}$ instead of A and $a_{i,j}$) yields

$$\det B = \underbrace{\binom{1-1}{1-1}}_{=1} \underbrace{\binom{2-1}{2-1}}_{=1} \cdots \underbrace{\binom{n-1}{n-1}}_{=1} = 1 \cdot 1 \cdot \cdots \cdot 1 = 1.$$

Exercise 6.4 (applied to B instead of A) shows that $\det\left(B^T\right) = \det B = 1$.

Now, we are going to show that $A = BB^T$.

Indeed, let us show that

$$\sum_{k=1}^{n} \binom{i-1}{k-1} \binom{j-1}{k-1} = \binom{i+j-2}{i-1} \tag{1058}$$

for every $(i,j) \in \{1,2,\ldots,n\}^2$.

[*Proof of (1058):* Let $(i,j) \in \{1,2,\ldots,n\}^2$. Then, $i \geq 1$, so that $i-1 \in \mathbb{N}$. Hence, Proposition 3.32 **(b)** (applied to $x = i-1$ and $y = j-1$) yields

$$\binom{(i-1)+(j-1)}{i-1} = \sum_{k=0}^{i-1} \binom{i-1}{k} \binom{j-1}{k}. \tag{1059}$$

On the other hand, for every $k \in \{i+1, i+2, \ldots, n\}$, we have

$$\binom{i-1}{k-1} = 0 \tag{1060}$$

[478]. Now, $i \leq n$, so that

$$\begin{aligned}
&\sum_{k=1}^{n} \binom{i-1}{k-1} \binom{j-1}{k-1} \\
&= \sum_{k=1}^{i} \binom{i-1}{k-1} \binom{j-1}{k-1} + \sum_{k=i+1}^{n} \underbrace{\binom{i-1}{k-1}}_{\substack{=0 \\ \text{(by (1060))}}} \binom{j-1}{k-1} \\
&= \sum_{k=1}^{i} \binom{i-1}{k-1} \binom{j-1}{k-1} + \underbrace{\sum_{k=i+1}^{n} 0 \binom{j-1}{k-1}}_{=0} = \sum_{k=1}^{i} \binom{i-1}{k-1} \binom{j-1}{k-1} \\
&= \sum_{k=0}^{i-1} \binom{i-1}{k} \binom{j-1}{k} \\
&\qquad \text{(here, we have substituted } k \text{ for } k-1 \text{ in the sum)} \\
&= \binom{(i-1)+(j-1)}{i-1} \qquad \text{(by (1059))} \\
&= \binom{i+j-2}{i-1} \qquad \left(\text{since } (i-1)+(j-1) = i+j-2\right).
\end{aligned}$$

This proves (1058).]

Now, we have $B = \left(\binom{i-1}{j-1} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ and therefore $B^T = \left(\binom{j-1}{i-1} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of B^T). Hence, the definition of the product BB^T shows that

$$BB^T = \left(\underbrace{\sum_{k=1}^{n} \binom{i-1}{k-1} \binom{j-1}{k-1}}_{\substack{= \binom{i+j-2}{i-1} \\ \text{(by (1058))}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(\binom{i+j-2}{i-1} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = A.$$

[478] *Proof of (1060):* Let $k \in \{i+1, i+2, \ldots, n\}$. Then, $k > i \geq 1$ and thus $k-1 \in \mathbb{N}$. Also, $k > i$, so that $k-i > i-1$ and thus $i-1 < k-1$. Hence, (231) (applied to $i-1$ and $k-1$ instead of m and n) yields $\binom{i-1}{k-1} = 0$, qed.

Hence, $A = BB^T$, so that

$$\det A = \det\left(BB^T\right) = \underbrace{\det B}_{=1} \cdot \underbrace{\det\left(B^T\right)}_{=1}$$
$$\left(\text{by Theorem 6.23, applied to } B \text{ and } B^T \text{ instead of } A \text{ and } B\right)$$
$$= 1.$$

This solves Exercise 6.11. □

7.79. Solution to Exercise 6.12

Proof of Lemma 6.34. **(b)** Let $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, m\}^n$ be an n-tuple satisfying $g_1 < g_2 < \cdots < g_n$. We shall derive a contradiction.

We have $m < n$, so that $n > m \geq 0$. Thus, $n \geq 1$ (since n is an integer). Therefore, the elements g_1 and g_n of $\{1, 2, \ldots, m\}$ are well-defined. Every $i \in \{1, 2, \ldots, n-1\}$ satisfies $g_i - i \leq g_{i+1} - (i+1)$ [479]. In other words, we have $g_1 - 1 \leq g_2 - 2 \leq \cdots \leq g_n - n$. Hence, $g_1 - 1 \leq g_n - n$.

But $g_n \in \{1, 2, \ldots, m\}$, so that $g_n \leq m < n$ and thus $\underbrace{g_n}_{<n} - n < n - n = 0$. Hence, $g_1 - 1 \leq g_n - n < 0$. On the other hand, $g_1 \in \{1, 2, \ldots, m\}$, so that $g_1 \geq 1$ and thus $g_1 - 1 \geq 0$. This contradicts $g_1 - 1 < 0$.

Now, let us forget that we fixed $(g_1, g_2, \ldots, g_n)$. We thus have derived a contradiction for every n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, m\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$. Hence, there exists no n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, m\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$. This proves Lemma 6.34 **(b)**.

(a) The n-tuple $(1, 2, \ldots, n)$ clearly belongs to $\{1, 2, \ldots, n\}^n$, and satisfies $1 < 2 < \cdots < n$. Hence, there exists an n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$, namely the n-tuple $(1, 2, \ldots, n)$. We shall now show that $(1, 2, \ldots, n)$ is the only such n-tuple.

Indeed, let $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ be an n-tuple satisfying $g_1 < g_2 < \cdots < g_n$. We shall prove that $(g_1, g_2, \ldots, g_n) = (1, 2, \ldots, n)$.

Every $i \in \{1, 2, \ldots, n-1\}$ satisfies $g_i - i \leq g_{i+1} - (i+1)$ [480]. In other words, we have $g_1 - 1 \leq g_2 - 2 \leq \cdots \leq g_n - n$. In other words,

$$g_u - u \leq g_v - v \tag{1061}$$

for any two elements u and v of $\{1, 2, \ldots, n\}$ satisfying $u \leq v$.

[479] *Proof.* Let $i \in \{1, 2, \ldots, n-1\}$. Then, $g_i < g_{i+1}$ (since $g_1 < g_2 < \cdots < g_n$). Hence, $g_i \leq g_{i+1} - 1$ (since g_i and g_{i+1} are integers). Hence, $\underbrace{g_i}_{\leq g_{i+1}-1} - i \leq g_{i+1} - 1 - i = g_{i+1} - (i+1)$, qed.

[480] *Proof.* Let $i \in \{1, 2, \ldots, n-1\}$. Then, $g_i < g_{i+1}$ (since $g_1 < g_2 < \cdots < g_n$). Hence, $g_i \leq g_{i+1} - 1$ (since g_i and g_{i+1} are integers). Hence, $\underbrace{g_i}_{\leq g_{i+1}-1} - i \leq g_{i+1} - 1 - i = g_{i+1} - (i+1)$, qed.

Let $i \in \{1, 2, \ldots, n\}$. Then, $1 \leq i \leq n$, so that $1 \leq n$. Hence, the elements g_1 and g_n of $\{1, 2, \ldots, n\}$ are well-defined.

Now, $1 \leq i$. Hence, (1061) (applied to $u = 1$ and $v = i$) yields $g_1 - 1 \leq g_i - i$. But $g_1 \in \{1, 2, \ldots, n\}$, so that $g_1 \geq 1$ and thus $g_1 - 1 \geq 0$. Hence, $0 \leq g_1 - 1 \leq g_i - i$.

On the other hand, $i \leq n$. Therefore, (1061) (applied to $u = i$ and $v = n$) yields $g_i - i \leq g_n - n$. But $g_n \in \{1, 2, \ldots, n\}$, so that $g_n \leq n$ and thus $g_n - n \leq 0$. Hence, $g_i - i \leq g_n - n \leq 0$. Combined with $0 \leq g_i - i$, this yields $g_i - i = 0$, so that $g_i = i$.

Now, let us forget that we fixed i. We thus have shown that $g_i = i$ for every $i \in \{1, 2, \ldots, n\}$. In other words, $(g_1, g_2, \ldots, g_n) = (1, 2, \ldots, n)$.

Let us now forget that we fixed $(g_1, g_2, \ldots, g_n)$. We thus have shown that if $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ is an n-tuple satisfying $g_1 < g_2 < \cdots < g_n$, then $(g_1, g_2, \ldots, g_n) = (1, 2, \ldots, n)$. In other words, every n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ must be equal to $(1, 2, \ldots, n)$. Hence, there exists at most one n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ (namely, the n-tuple $(1, 2, \ldots, n)$).

We now know the following two facts:

- There exists an n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$, namely the n-tuple $(1, 2, \ldots, n)$.
- There exists at most one n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$.

Combining these two facts, we conclude that there exists exactly one n-tuple $(g_1, g_2, \ldots, g_n) \in \{1, 2, \ldots, n\}^n$ satisfying $g_1 < g_2 < \cdots < g_n$, namely the n-tuple $(1, 2, \ldots, n)$. This proves Lemma 6.34 **(a)**. □

7.80. Solution to Exercise 6.13

Before we give a formal proof of Proposition 6.40, let us outline the main ideas of this proof: We shall define the notion of *inversion* of an n-tuple $(a_1, a_2, \ldots, a_n)$ of integers[481]; then we will argue that switching two adjacent entries a_k and a_{k+1} of an n-tuple $(a_1, a_2, \ldots, a_n)$ which are "out of order" (i.e., satisfy $a_k > a_{k+1}$) reduces the number of inversions by 1 (this is our equality (1072) further below), and thus a sequence of such switches will eventually end and therefore bring the tuple into weakly increasing order. (This is similar to an argument in the solution of Exercise 5.2 **(e)**.) Other proofs of Proposition 6.40 **(a)** are possible[482]. Proposition 6.40 **(b)** will then easily follow (indeed, we will argue that $a_{\sigma(i)}$ is the smallest integer x such that at least i different elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq x$), and Proposition 6.40 **(c)** will finally follow from parts **(a)** and **(b)**.

[481] It will be defined as a pair (i, j) of integers satisfying $1 \leq i < j \leq n$ and $a_i > a_j$. The analogy with the notion of "inversion" of a permutation is intentional.

[482] Roughly speaking, to each sorting algorithm corresponds at least one proof of Proposition 6.40 **(a)**. ("At least" because there are often several ways to prove the correctness of a given sorting algorithm.) The proof we just outlined corresponds to "bubble sort".

Here is the proof in detail:

Proof of Proposition 6.40. **(a)** Let us forget that we fixed $a_1, a_2, \ldots, a_n$.

We shall first introduce some notations.

We let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is an n-tuple of integers, then:

- An *inversion* of $\mathbf{a}$ will mean a pair $(i, j) \in [n]^2$ satisfying $i < j$ and $a_i > a_j$.
- We denote by $\operatorname{Inv}(\mathbf{a})$ the set of all inversions of $\mathbf{a}$. Thus, $\operatorname{Inv}(\mathbf{a}) \subseteq [n]^2$. More precisely,

$$\begin{aligned} \operatorname{Inv}(\mathbf{a}) &= (\text{the set of all inversions of } \mathbf{a}) \\ &= \left\{ (i, j) \in [n]^2 \mid i < j \text{ and } a_i > a_j \right\} \qquad (1062) \\ &\qquad \left(\begin{array}{c} \text{since the inversions of } \mathbf{a} \text{ are the pairs } (i, j) \in [n]^2 \text{ satisfying} \\ i < j \text{ and } a_i > a_j \text{ (by the definition of an "inversion")} \end{array} \right) \\ &= \left\{ (u, v) \in [n]^2 \mid u < v \text{ and } a_u > a_v \right\} \qquad (1063) \\ &\qquad (\text{here, we renamed the index } (i, j) \text{ as } (u, v)). \end{aligned}$$

- We denote by $\ell(\mathbf{a})$ the number $|\operatorname{Inv}(\mathbf{a})|$. (This is well-defined because $\operatorname{Inv}(\mathbf{a})$ is finite (since $\operatorname{Inv}(\mathbf{a}) \subseteq [n]^2$).) Thus,

$$\begin{aligned} \ell(\mathbf{a}) &= \left| \underbrace{\operatorname{Inv}(\mathbf{a})}_{=(\text{the set of all inversions of } \mathbf{a})} \right| = |(\text{the set of all inversions of } \mathbf{a})| \\ &= (\text{the number of all inversions of } \mathbf{a}). \qquad (1064) \end{aligned}$$

- For every permutation $\tau \in S_n$, we denote by $\mathbf{a} \circ \tau$ the n-tuple $\left(a_{\tau(1)}, a_{\tau(2)}, \ldots, a_{\tau(n)} \right)$ of integers.

We notice that if $\mathbf{a}$ is any n-tuple of integers, then

$$\mathbf{a} \circ (\sigma \circ \tau) = (\mathbf{a} \circ \sigma) \circ \tau \qquad \text{for any } \sigma \in S_n \text{ and } \tau \in S_n \qquad (1065)$$

[483]. Also, if $\mathbf{a}$ is any n-tuple of integers, then

$$\mathbf{a} \circ \operatorname{id} = \mathbf{a} \qquad (1066)$$

[483] *Proof of (1065):* Let $\mathbf{a}$ be any n-tuple of integers. Let $\sigma \in S_n$ and $\tau \in S_n$. We must show that $\mathbf{a} \circ (\sigma \circ \tau) = (\mathbf{a} \circ \sigma) \circ \tau$.

Write the n-tuple $\mathbf{a}$ in the form $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ for some integers $a_1, a_2, \ldots, a_n$. Thus, the definition of $\mathbf{a} \circ (\sigma \circ \tau)$ yields

$$\mathbf{a} \circ (\sigma \circ \tau) = \left(a_{(\sigma \circ \tau)(1)}, a_{(\sigma \circ \tau)(2)}, \ldots, a_{(\sigma \circ \tau)(n)} \right) = \left(a_{\sigma(\tau(1))}, a_{\sigma(\tau(2))}, \ldots, a_{\sigma(\tau(n))} \right)$$

[484].

If X, X', Y and Y' are four sets and if $\alpha : X \to X'$ and $\beta : Y \to Y'$ are two maps, then $\alpha \times \beta$ will denote the map

$$X \times Y \to X' \times Y',$$
$$(x, y) \mapsto (\alpha(x), \beta(y)).$$

Recall that, for each $k \in \{1, 2, \ldots, n-1\}$, we have defined s_k to be the permutation in S_n that switches k with $k+1$ but leaves all other numbers unchanged. This permutation s_k is a map $[n] \to [n]$ and satisfies $s_k^2 = \text{id}$. For every $k \in \{1, 2, \ldots, n-1\}$, the map $s_k \times s_k : [n] \times [n] \to [n] \times [n]$ satisfies $(s_k \times s_k)^2 = \text{id}$ [485], and thus is a bijection from $[n]^2$ to $[n]^2$ [486].

(since $a_{(\sigma\circ\tau)(i)} = a_{\sigma(\tau(i))}$ for every $i \in \{1, 2, \ldots, n\}$). On the other hand, the definition of $\mathbf{a} \circ \sigma$ yields $\mathbf{a} \circ \sigma = \left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right)$ (since $\mathbf{a} = (a_1, a_2, \ldots, a_n)$). Therefore, the definition of $(\mathbf{a} \circ \sigma) \circ \tau$ yields $(\mathbf{a} \circ \sigma) \circ \tau = \left(a_{\sigma(\tau(1))}, a_{\sigma(\tau(2))}, \ldots, a_{\sigma(\tau(n))}\right)$. Compared with $\mathbf{a} \circ (\sigma \circ \tau) = \left(a_{\sigma(\tau(1))}, a_{\sigma(\tau(2))}, \ldots, a_{\sigma(\tau(n))}\right)$, this yields $\mathbf{a} \circ (\sigma \circ \tau) = (\mathbf{a} \circ \sigma) \circ \tau$. This proves (1065).

[484] *Proof of (1066):* Let $\mathbf{a}$ be any n-tuple of integers.

Write the n-tuple $\mathbf{a}$ in the form $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ for some integers $a_1, a_2, \ldots, a_n$. Thus, the definition of $\mathbf{a} \circ \text{id}$ yields

$$\mathbf{a} \circ \text{id} = \left(a_{\text{id}(1)}, a_{\text{id}(2)}, \ldots, a_{\text{id}(n)}\right) = (a_1, a_2, \ldots, a_n)$$

(since $a_{\text{id}(i)} = a_i$ for every $i \in \{1, 2, \ldots, n\}$). Compared with $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, this yields $\mathbf{a} \circ \text{id} = \mathbf{a}$. This proves (1066).

[485] *Proof.* Let $u \in [n] \times [n]$. Then, we can write u in the form (i, j) for some $i \in [n]$ and $j \in [n]$. Consider these i and j. We have

$$\underbrace{(s_k \times s_k)^2}_{=(s_k\times s_k)\circ(s_k\times s_k)} \Big(\underbrace{u}_{=(i,j)}\Big) = ((s_k \times s_k) \circ (s_k \times s_k))((i, j)) = (s_k \times s_k) \Bigg(\underbrace{(s_k \times s_k)((i, j))}_{\substack{=(s_k(i), s_k(j)) \\ \text{(by the definition of } s_k\times s_k)}}\Bigg)$$
$$= (s_k \times s_k)((s_k(i), s_k(j))) = \Big(\underbrace{s_k(s_k(i))}_{=s_k^2(i)}, \underbrace{s_k(s_k(j))}_{=s_k^2(j)}\Big)$$
$$\text{(by the definition of } s_k \times s_k)$$
$$= \Big(\underbrace{s_k^2}_{=\text{id}}(i), \underbrace{s_k^2}_{=\text{id}}(j)\Big) = \Big(\underbrace{\text{id}(i)}_{=i}, \underbrace{\text{id}(j)}_{=j}\Big) = (i, j) = u = \text{id}(u).$$

Now, let us forget that we fixed u. We thus have shown that $(s_k \times s_k)^2(u) = \text{id}(u)$ for every $u \in [n] \times [n]$. In other words, $(s_k \times s_k)^2 = \text{id}$, qed.

[486] *Proof.* Let $k \in \{1, 2, \ldots, n-1\}$. We have $(s_k \times s_k) \circ (s_k \times s_k) = (s_k \times s_k)^2 = \text{id}$. Hence, the maps $s_k \times s_k$ and $s_k \times s_k$ are mutually inverse. Hence, the map $s_k \times s_k$ is invertible, thus a bijection.

Let us now recall a simple fact: If u and v are two integers such that $1 \leq u < v \leq n$, and if $k \in \{1, 2, \ldots, n-1\}$ is such that $(u, v) \neq (k, k+1)$, then

$$s_k(u) < s_k(v). \tag{1067}$$

(This was proven in the solution of Exercise 5.2 **(a)**.)

Now, we notice the following facts:

- If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is an n-tuple of integers, and if $k \in \{1, 2, \ldots, n-1\}$, then

$$\operatorname{Inv}(\mathbf{a} \circ s_k) = \left\{ (u, v) \in [n]^2 \mid u < v \text{ and } a_{s_k(u)} > a_{s_k(v)} \right\} \tag{1068}$$

[487].

- If $\mathbf{a}$ is an n-tuple of integers, and if $k \in \{1, 2, \ldots, n-1\}$, then

$$(s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}) \subseteq \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\} \tag{1069}$$

[488].

Therefore, this map $s_k \times s_k$ is a bijection from $[n]^2$ to $[n]^2$ (because its domain is $[n] \times [n] = [n]^2$, and its codomain is $[n] \times [n] = [n]^2$). Qed.

[487] *Proof of (1068):* Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-tuple of integers, and let $k \in \{1, 2, \ldots, n-1\}$. The definition of $\mathbf{a} \circ s_k$ yields $\mathbf{a} \circ s_k = \left(a_{s_k(1)}, a_{s_k(2)}, \ldots, a_{s_k(n)} \right)$ (since $\mathbf{a} = (a_1, a_2, \ldots, a_n)$). Hence, (1063) (applied to $\mathbf{a} \circ s_k$ and $a_{s_k(i)}$ instead of $\mathbf{a}$ and a_i) yields $\operatorname{Inv}(\mathbf{a} \circ s_k) = \left\{ (u, v) \in [n]^2 \mid u < v \text{ and } a_{s_k(u)} > a_{s_k(v)} \right\}$. This proves (1068).

[488] *Proof of (1069):* Let $\mathbf{a}$ be an n-tuple of integers, and let $k \in \{1, 2, \ldots, n-1\}$. Write the n-tuple $\mathbf{a}$ in the form $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ for some integers $a_1, a_2, \ldots, a_n$.

Let $c \in (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\})$. Thus, $c \in [n]^2$, so that we can write c in the form $c = (i, j)$ for some $i \in [n]$ and $j \in [n]$. Consider these i and j.

We have $(i, j) = c \in (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\})$, so that $(s_k \times s_k)((i, j)) \in \operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}$. Since $(s_k \times s_k)((i, j)) = (s_k(i), s_k(j))$ (by the definition of $(s_k \times s_k)$), this rewrites as $(s_k(i), s_k(j)) \in \operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}$. In other words, $(s_k(i), s_k(j)) \in \operatorname{Inv}(\mathbf{a})$ and $(s_k(i), s_k(j)) \neq (k, k+1)$.

We have $(s_k(i), s_k(j)) \in \operatorname{Inv}(\mathbf{a}) = \left\{ (u, v) \in [n]^2 \mid u < v \text{ and } a_u > a_v \right\}$ (by (1063)). In other words, $(s_k(i), s_k(j))$ is an element of $[n]^2$ and satisfies $s_k(i) < s_k(j)$ and $a_{s_k(i)} > a_{s_k(j)}$. We have $s_k(i) < s_k(j)$ and $(s_k(i), s_k(j)) \neq (k, k+1)$. Thus, we can apply (1067) to $u = s_k(i)$ and $v = s_k(j)$. We thus conclude $s_k(s_k(i)) < s_k(s_k(j))$. In other words, $i < j$ (since $s_k(s_k(i)) = \underbrace{s_k^2}_{=\operatorname{id}}(i) = \operatorname{id}(i) = i$ and $s_k(s_k(j)) = \underbrace{s_k^2}_{=\operatorname{id}}(j) = \operatorname{id}(j) = j$).

Now, we know that (i, j) is an element (u, v) of $[n]^2$ satisfying $u < v$ and $a_{s_k(u)} > a_{s_k(v)}$ (since $i < j$ and $a_{s_k(i)} > a_{s_k(j)}$). In other words,

$$(i, j) \in \left\{ (u, v) \in [n]^2 \mid u < v \text{ and } a_{s_k(u)} > a_{s_k(v)} \right\} = \operatorname{Inv}(\mathbf{a} \circ s_k)$$

(by (1068)).

Next, let us assume (for the sake of contradiction) that $(i, j) = (k, k+1)$. Thus, $i = k$ and

- If $\mathbf{a}$ is an n-tuple of integers, and if $k \in \{1, 2, \ldots, n-1\}$, then

$$\operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\} \subseteq (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}) \qquad (1070)$$

[489].

$j = k+1$. Thus, $s_k\left(\underbrace{i}_{=k}\right) = s_k(k) = k+1 > k = s_k\left(\underbrace{k+1}_{=j}\right) = s_k(j)$; but this contradicts $s_k(i) < s_k(j)$. This contradiction shows that our assumption (that $(i, j) = (k, k+1)$) was wrong. Hence, we cannot have $(i, j) = (k, k+1)$. In other words, we must have $(i, j) \neq (k, k+1)$. Combined with $(i, j) \in \operatorname{Inv}(\mathbf{a} \circ s_k)$, this yields $(i, j) \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}$. Thus, $c = (i, j) \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}$.

Now, let us forget that we fixed c. We thus have shown that $c \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}$ for every $c \in (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\})$. In other words,

$$(s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}) \subseteq \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}.$$

This proves (1069).

[489]*Proof of (1070):* Let $\mathbf{a}$ be an n-tuple of integers, and let $k \in \{1, 2, \ldots, n-1\}$. Write the n-tuple $\mathbf{a}$ in the form $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ for some integers $a_1, a_2, \ldots, a_n$.

Let $c \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}$. Thus, $c \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\} \subseteq \operatorname{Inv}(\mathbf{a} \circ s_k) \subseteq [n]^2$, so that we can write c in the form $c = (i, j)$ for some $i \in [n]$ and $j \in [n]$. Consider these i and j.

We have $(i, j) = c \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}$. In other words, $(i, j) \in \operatorname{Inv}(\mathbf{a} \circ s_k)$ and $(i, j) \neq (k, k+1)$.

We have $(i, j) \in \operatorname{Inv}(\mathbf{a} \circ s_k) = \left\{(u, v) \in [n]^2 \mid u < v \text{ and } a_{s_k(u)} > a_{s_k(v)}\right\}$ (by (1068)). In other words, (i, j) is an element of $[n]^2$ and satisfies $i < j$ and $a_{s_k(i)} > a_{s_k(j)}$. Applying (1067) to $u = i$ and $v = j$, we obtain $s_k(i) < s_k(j)$ (since $i < j$ and $(i, j) \neq (k, k+1)$).

The pair $(s_k(i), s_k(j))$ is a pair $(u, v) \in [n]^2$ satisfying $u < v$ and $a_u > a_v$ (since $s_k(i) < s_k(j)$ and $a_{s_k(i)} > a_{s_k(j)}$). In other words, $(s_k(i), s_k(j)) \in \left\{(u, v) \in [n]^2 \mid u < v \text{ and } a_u > a_v\right\} = \operatorname{Inv}(\mathbf{a})$ (by (1063)).

Next, let us assume (for the sake of contradiction) that $(s_k(i), s_k(j)) = (k, k+1)$. Thus, $s_k(i) = k$ and $s_k(j) = k+1$. Hence, $k+1 = s_k\left(\underbrace{k}_{=s_k(i)}\right) = s_k(s_k(i)) = \underbrace{s_k^2}_{=\operatorname{id}}(i) = i$ and $k = s_k\left(\underbrace{k+1}_{=s_k(j)}\right) = s_k(s_k(j)) = \underbrace{s_k^2}_{=\operatorname{id}}(j) = \operatorname{id}(j) = j$, so that $i = k+1 > k = j$. This contradicts $i < j$. This contradiction shows that our assumption (that $(s_k(i), s_k(j)) = (k, k+1)$) was wrong. Hence, we cannot have $(s_k(i), s_k(j)) = (k, k+1)$. In other words, we must have $(s_k(i), s_k(j)) \neq (k, k+1)$. Combined with $(s_k(i), s_k(j)) \in \operatorname{Inv}(\mathbf{a})$, this yields $(s_k(i), s_k(j)) \in \operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}$.

The definition of $s_k \times s_k$ yields $(s_k \times s_k)((i, j)) = (s_k(i), s_k(j)) \in \operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}$. In other words, $(i, j) \in (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\})$. Hence,

$$c = (i, j) \in (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}).$$

Now, let us forget that we fixed c. We thus have shown that $c \in (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\})$ for every $c \in \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\}$. In other words,

$$\operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\} \subseteq (s_k \times s_k)^{-1} (\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}).$$

This proves (1070).

- If $\mathbf{a}$ is an n-tuple of integers, and if $k \in \{1, 2, \ldots, n-1\}$, then

$$\left(s_k \times s_k\right)^{-1} \left(\operatorname{Inv}\left(\mathbf{a}\right) \setminus \left\{\left(k, k+1\right)\right\}\right) = \operatorname{Inv}\left(\mathbf{a} \circ s_k\right) \setminus \left\{\left(k, k+1\right)\right\} \tag{1071}$$

[490].

- If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is an n-tuple of integers, and if $k \in \{1, 2, \ldots, n-1\}$ is such that $a_k > a_{k+1}$, then

$$\ell\left(\mathbf{a} \circ s_k\right) = \ell\left(\mathbf{a}\right) - 1 \tag{1072}$$

[491].

- If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is an n-tuple of integers satisfying $\ell(\mathbf{a}) = 0$, then

$$a_1 \leq a_2 \leq \cdots \leq a_n \tag{1074}$$

[492].

[490] *Proof of (1071):* Let $\mathbf{a}$ be an n-tuple of integers, and let $k \in \{1, 2, \ldots, n-1\}$. Then, combining (1069) with (1070), we obtain $\left(s_k \times s_k\right)^{-1} \left(\operatorname{Inv}\left(\mathbf{a}\right) \setminus \left\{\left(k, k+1\right)\right\}\right) = \operatorname{Inv}\left(\mathbf{a} \circ s_k\right) \setminus \left\{\left(k, k+1\right)\right\}$. This proves (1071).

[491] *Proof of (1072):* Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-tuple of integers, and let $k \in \{1, 2, \ldots, n-1\}$ be such that $a_k > a_{k+1}$.

The definition of $\ell(\mathbf{a})$ yields $\ell(\mathbf{a}) = |\operatorname{Inv}(\mathbf{a})|$.

The pair $(k, k+1)$ is a pair $(u, v) \in [n]^2$ such that $u < v$ and $a_u > a_v$ (since $k < k+1$ and $a_k > a_{k+1}$). In other words,

$$(k, k+1) \in \left\{(u, v) \in [n]^2 \mid u < v \text{ and } a_u > a_v\right\} = \operatorname{Inv}(\mathbf{a}) \qquad \text{(by (1063))}.$$

Hence, $|\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}| = \underbrace{|\operatorname{Inv}(\mathbf{a})|}_{=\ell(\mathbf{a})} - 1 = \ell(\mathbf{a}) - 1$. The map $s_k \times s_k$ is a bijection, and thus we have $\left|\left(s_k \times s_k\right)^{-1}(X)\right| = |X|$ for every subset X of $[n]^2$. Applying this to $X = \operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}$, we obtain

$$\left|\left(s_k \times s_k\right)^{-1} \left(\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}\right)\right| = |\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}| = \ell(\mathbf{a}) - 1. \tag{1073}$$

On the other hand, let us assume (for the sake of contradiction) that $(k, k+1) \in \operatorname{Inv}(\mathbf{a} \circ s_k)$. Thus,

$$(k, k+1) \in \operatorname{Inv}(\mathbf{a} \circ s_k) = \left\{(u, v) \in [n]^2 \mid u < v \text{ and } a_{s_k(u)} > a_{s_k(v)}\right\}$$

(by (1068)). In other words, $(k, k+1)$ is a pair $(u, v) \in [n]^2$ such that $u < v$ and $a_{s_k(u)} > a_{s_k(v)}$. In other words, $k < k+1$ and $a_{s_k(k)} > a_{s_k(k+1)}$. Now, $a_{s_k(k)} > a_{s_k(k+1)}$. In other words, $a_{k+1} > a_k$ (since $s_k(k) = k+1$ and $s_k(k+1) = k$). This contradicts $a_k > a_{k+1}$. This contradiction shows that our assumption (that $(k, k+1) \in \operatorname{Inv}(\mathbf{a} \circ s_k)$) was wrong. Hence, we cannot have $(k, k+1) \in \operatorname{Inv}(\mathbf{a} \circ s_k)$. We thus have $(k, k+1) \notin \operatorname{Inv}(\mathbf{a} \circ s_k)$. Hence, $\operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\} = \operatorname{Inv}(\mathbf{a} \circ s_k)$. Now, (1071) becomes

$$\left(s_k \times s_k\right)^{-1} \left(\operatorname{Inv}(\mathbf{a}) \setminus \{(k, k+1)\}\right) = \operatorname{Inv}(\mathbf{a} \circ s_k) \setminus \{(k, k+1)\} = \operatorname{Inv}(\mathbf{a} \circ s_k).$$

Therefore, (1073) rewrites as $|\operatorname{Inv}(\mathbf{a} \circ s_k)| = \ell(\mathbf{a}) - 1$.

But the definition of $\ell(\mathbf{a} \circ s_k)$ yields $\ell(\mathbf{a} \circ s_k) = |\operatorname{Inv}(\mathbf{a} \circ s_k)| = \ell(\mathbf{a}) - 1$. This proves (1072).

[492] *Proof of (1074):* Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-tuple of integers such that $\ell(\mathbf{a}) = 0$.

- If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is an n-tuple of integers satisfying $a_1 \leq a_2 \leq \cdots \leq a_n$, then
$$\ell(\mathbf{a}) = 0 \tag{1075}$$
[493].

- If $\mathbf{a}$ is an n-tuple of integers, then
$$\text{there exists a } \sigma \in S_n \text{ such that } \ell(\mathbf{a} \circ \sigma) = 0 \tag{1077}$$
[494].

We assume (for the sake of contradiction) that there exists some $k \in \{1, 2, \ldots, n-1\}$ such that $a_k > a_{k+1}$. Consider this k. Then, $(k, k+1)$ is an element $(u, v) \in [n]^2$ satisfying $u < v$ and $a_u > a_v$ (since $k < k+1$ and $a_k > a_{k+1}$). In other words,
$$(k, k+1) \in \left\{(u, v) \in [n]^2 \mid u < v \text{ and } a_u > a_v\right\} = \operatorname{Inv}(\mathbf{a})$$
(by (1063)). Hence, the set $\operatorname{Inv}(\mathbf{a})$ contains at least one element (namely, $(k, k+1)$). In other words, $|\operatorname{Inv}(\mathbf{a})| \geq 1$.

But the definition of $\ell(\mathbf{a})$ yields $\ell(\mathbf{a}) = |\operatorname{Inv}(\mathbf{a})| \geq 1$. This contradicts $\ell(\mathbf{a}) = 0$. This contradiction shows that our assumption (that there exists some $k \in \{1, 2, \ldots, n-1\}$ such that $a_k > a_{k+1}$) was wrong.

Therefore, there exists no $k \in \{1, 2, \ldots, n-1\}$ such that $a_k > a_{k+1}$. In other words, every $k \in \{1, 2, \ldots, n-1\}$ satisfies $a_k \leq a_{k+1}$. In other words, $a_1 \leq a_2 \leq \cdots \leq a_n$. This proves (1074).

[493] *Proof of (1075):* Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be an n-tuple of integers satisfying $a_1 \leq a_2 \leq \cdots \leq a_n$. We must prove that $\ell(\mathbf{a}) = 0$.

Indeed, assume the contrary. Thus, $\ell(\mathbf{a}) \neq 0$. But the definition of $\ell(\mathbf{a})$ yields $\ell(\mathbf{a}) = |\operatorname{Inv}(\mathbf{a})|$, so that $|\operatorname{Inv}(\mathbf{a})| = \ell(\mathbf{a}) \neq 0$. Hence, the set $\operatorname{Inv}(\mathbf{a})$ is nonempty. In other words, there exists some $c \in \operatorname{Inv}(\mathbf{a})$. Consider this c.

Recall that $a_1 \leq a_2 \leq \cdots \leq a_n$. In other words,
$$a_u \leq a_v \tag{1076}$$
for any $u \in [n]$ and $v \in [n]$ satisfying $u < v$.

But
$$c \in \operatorname{Inv}(\mathbf{a}) = \left\{(u, v) \in [n]^2 \mid u < v \text{ and } a_u > a_v\right\}$$
(by (1063)). In other words, c has the form $c = (u, v)$ for some $(u, v) \in [n]^2$ satisfying $u < v$ and $a_u > a_v$. Consider this $(u, v) \in [n]^2$. We have $u < v$, and thus $a_u \leq a_v$ (by (1076)). This contradicts $a_u > a_v$. This contradiction proves that our assumption was wrong.

Hence, we have shown that $\ell(\mathbf{a}) = 0$. This proves (1075).

[494] *Proof of (1077):* We shall prove (1077) by induction over $\ell(\mathbf{a})$:

Induction base: If $\mathbf{a}$ is an n-tuple of integers satisfying $\ell(\mathbf{a}) = 0$, then we have $\ell\left(\underbrace{\mathbf{a} \circ \operatorname{id}}_{\substack{=\mathbf{a} \\ \text{(by (1066))}}}\right) = \ell(\mathbf{a}) = 0$. Hence, if $\mathbf{a}$ is an n-tuple of integers satisfying $\ell(\mathbf{a}) = 0$, then there exists a $\sigma \in S_n$ such that $\ell(\mathbf{a} \circ \sigma) = 0$ (namely, $\sigma = \operatorname{id}$). In other words, (1077) holds for $\ell(\mathbf{a}) = 0$. This completes the induction base.

Induction step: Let L be a positive integer. Assume that (1077) holds for $\ell(\mathbf{a}) = L - 1$. We need to prove that (1077) holds for $\ell(\mathbf{a}) = L$.

Now, let us prove Proposition 6.40 **(a)**. Let $a_1, a_2, \ldots, a_n$ be n integers. Let $\mathbf{a}$ be the n-tuple $(a_1, a_2, \ldots, a_n)$. Then, there exists a $\sigma \in S_n$ such that $\ell(\mathbf{a} \circ \sigma) = 0$ (by (1077)). Consider this σ. The definition of $\mathbf{a} \circ \sigma$ shows that $\mathbf{a} \circ \sigma = \left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right)$ (since $\mathbf{a} = (a_1, a_2, \ldots, a_n)$). Therefore, (1074) (applied to $\mathbf{a} \circ \sigma$ and $a_{\sigma(i)}$ instead of $\mathbf{a}$ and a_i) yields $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$.

Let us now forget that we defined σ. We thus have constructed a $\sigma \in S_n$ satisfying $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. Therefore, there exists a permutation $\sigma \in S_n$ such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. This proves Proposition 6.40 **(a)**.

(b) Let $\sigma \in S_n$ be such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. Let $i \in \{1, 2, \ldots, n\}$. We shall now show that

$$a_{\sigma(i)} = \min\left\{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\right\}. \qquad (1079)$$

(This statement includes the tacit claim that the right hand side of (1079) is well-defined.)

[*Proof of (1079):* Define a subset X of $\mathbb{Z}$ by

$$X = \left\{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\right\}.$$

We shall show that $a_{\sigma(i)} = \min X$.

We have $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. In other words,

$$a_{\sigma(u)} \leq a_{\sigma(v)} \qquad (1080)$$

We have assumed that (1077) holds for $\ell(\mathbf{a}) = L - 1$. In other words, if $\mathbf{a}$ is an n-tuple of integers satisfying $\ell(\mathbf{a}) = L - 1$, then

$$\text{there exists a } \sigma \in S_n \text{ such that } \ell(\mathbf{a} \circ \sigma) = 0. \qquad (1078)$$

Now, let $\mathbf{a}$ be an n-tuple of integers satisfying $\ell(\mathbf{a}) = L$. We shall show that there exists a $\sigma \in S_n$ such that $\ell(\mathbf{a} \circ \sigma) = 0$.

Let us first prove that there exists some $k \in \{1, 2, \ldots, n-1\}$ such that $a_k > a_{k+1}$. In fact, assume the contrary. Thus, there exists no $k \in \{1, 2, \ldots, n-1\}$ such that $a_k > a_{k+1}$. In other words, every $k \in \{1, 2, \ldots, n-1\}$ satisfies $a_k \leq a_{k+1}$. In other words, $a_1 \leq a_2 \leq \cdots \leq a_n$. Thus, (1075) shows that $\ell(\mathbf{a}) = 0$. Hence, $0 = \ell(\mathbf{a}) = L > 0$ (since L is positive). This is absurd. This contradiction proves that our assumption was wrong.

Hence, we have proven that there exists some $k \in \{1, 2, \ldots, n-1\}$ such that $a_k > a_{k+1}$. Consider this k. Then, (1072) shows that $\ell(\mathbf{a} \circ s_k) = \underbrace{\ell(\mathbf{a})}_{=L} - 1 = L - 1$. Therefore, (1078) (applied to $\mathbf{a} \circ s_k$ instead of $\mathbf{a}$) shows that there exists a $\sigma \in S_n$ such that $\ell((\mathbf{a} \circ s_k) \circ \sigma) = 0$. Let τ be such a σ. Thus, τ is an element of S_n and satisfies $\ell((\mathbf{a} \circ s_k) \circ \tau) = 0$.

Now, (1065) (applied to s_k instead of σ) yields $\mathbf{a} \circ (s_k \circ \tau) = (\mathbf{a} \circ s_k) \circ \tau$. Hence, $\ell(\mathbf{a} \circ (s_k \circ \tau)) = \ell((\mathbf{a} \circ s_k) \circ \tau) = 0$. Thus, there exists a $\sigma \in S_n$ such that $\ell(\mathbf{a} \circ \sigma) = 0$ (namely, $\sigma = s_k \circ \tau$).

Now, let us forget that we fixed $\mathbf{a}$. We thus have shown that if $\mathbf{a}$ is an n-tuple of integers satisfying $\ell(\mathbf{a}) = L$, then there exists a $\sigma \in S_n$ such that $\ell(\mathbf{a} \circ \sigma) = 0$. In other words, (1077) holds for $\ell(\mathbf{a}) = L$. This completes the induction step. Thus, the induction proof of (1077) is complete.

for any two elements u and v of $\{1, 2, \ldots, n\}$ satisfying $u \leq v$.

We have $\sigma \in S_n$. Thus, σ is a permutation, thus a bijective map, thus an injective map.

Every $k \in \{1, 2, \ldots, i\}$ satisfies $\sigma(k) \in \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq a_{\sigma(i)} \right\}$ [495]. In other words,

$$\{\sigma(k) \mid k \in \{1, 2, \ldots, i\}\} \subseteq \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq a_{\sigma(i)} \right\}.$$

Hence,

$$|\{\sigma(k) \mid k \in \{1, 2, \ldots, i\}\}| \leq \left| \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq a_{\sigma(i)} \right\} \right|,$$

so that

$$\begin{aligned} \left| \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq a_{\sigma(i)} \right\} \right| &\geq \left| \underbrace{\{\sigma(k) \mid k \in \{1, 2, \ldots, i\}\}}_{=\sigma(\{1,2,\ldots,i\})} \right| \\ &= |\sigma(\{1, 2, \ldots, i\})| = |\{1, 2, \ldots, i\}| \\ &\qquad (\text{since the map } \sigma \text{ is injective}) \\ &= i. \end{aligned}$$

In other words, at least i elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq a_{\sigma(i)}$. In other words, $a_{\sigma(i)}$ is an element x of $\mathbb{Z}$ such that at least i elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq x$. In other words,

$$a_{\sigma(i)} \in \left\{ x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x \right\} = X.$$

On the other hand, let y be any element of X. We shall show that $a_{\sigma(i)} \leq y$. Indeed, assume the contrary. Thus, $a_{\sigma(i)} > y$. Hence,

$$\left| \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq y \right\} \right| < i \tag{1081}$$

[496].

But

$$y \in X = \left\{ x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x \right\}.$$

[495] *Proof.* Let $k \in \{1, 2, \ldots, i\}$. Then, $k \leq i$ and thus $a_{\sigma(k)} \leq a_{\sigma(i)}$ (by (1080), applied to $u = k$ and $v = i$). Hence, $\sigma(k)$ is an element j of $\{1, 2, \ldots, n\}$ satisfying $a_j \leq a_{\sigma(i)}$ (because $a_{\sigma(k)} \leq a_{\sigma(i)}$). In other words, $\sigma(k) \in \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq a_{\sigma(i)} \right\}$, qed.

[496] *Proof.* Let $p \in \left\{ j \in \{1, 2, \ldots, n\} \mid a_j \leq y \right\}$. Thus, p is an element j of $\{1, 2, \ldots, n\}$ such that $a_j \leq y$. In other words, p is an element of $\{1, 2, \ldots, n\}$ and satisfies $a_p \leq y$.

The permutation σ has an inverse σ^{-1}. Let $q = \sigma^{-1}(p)$. Thus, $p = \sigma(q)$. Hence, $a_p = a_{\sigma(q)}$, so that $a_{\sigma(q)} = a_p \leq y < a_{\sigma(i)}$ (since $a_{\sigma(i)} > y$).

If we had $i \leq q$, then we would have $a_{\sigma(i)} \leq a_{\sigma(q)}$ (by (1080), applied to $u = i$ and $v = q$), which would contradict $a_{\sigma(q)} < a_{\sigma(i)}$. Hence, we cannot have $i \leq q$. Thus, we must have $q < i$.

In other words, y is an element x of $\mathbb{Z}$ such that at least i elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq x$. In other words, y is an element of $\mathbb{Z}$, and at least i elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq y$. We have

$$\left|\left\{j \in \{1, 2, \ldots, n\} \ \mid \ a_j \leq y\right\}\right| \geq i$$

(since at least i elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq y$). This contradicts (1081). This contradiction proves that our assumption was wrong. Hence, $a_{\sigma(i)} \leq y$ is proven.

Now, let us forget that we fixed y. We thus have shown that every $y \in X$ satisfies $a_{\sigma(i)} \leq y$. Altogether, we thus have shown the following two facts:

- We have $a_{\sigma(i)} \in X$.
- Every $y \in X$ satisfies $a_{\sigma(i)} \leq y$.

Combining these two facts, we obtain $a_{\sigma(i)} = \min X$ (and, in particular, this shows that $\min X$ is well-defined). Since $X = \left\{x \in \mathbb{Z} \ \mid \ \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\right\}$, this rewrites as follows:

$$a_{\sigma(i)} = \min\left\{x \in \mathbb{Z} \ \mid \ \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\right\}.$$

Thus, (1079) is proven.]

Now, the value $a_{\sigma(i)}$ depends only on $a_1, a_2, \ldots, a_n$ and i (but not on σ) (because (1079) gives a description of $a_{\sigma(i)}$ which involves $a_1, a_2, \ldots, a_n$ and i, but not σ). Thus, Proposition 6.40 **(b)** is proven.

(c) The integers $a_1, a_2, \ldots, a_n$ are distinct. In other words, if u and v are two distinct elements of $\{1, 2, \ldots, n\}$, then

$$a_u \neq a_v. \tag{1082}$$

We have $\sigma \in S_n$. Thus, σ is a permutation, thus a bijective map, thus an injective map.

Now, we can see that:

Hence, $q \in \{1, 2, \ldots, i-1\}$. Thus, $p = \sigma\left(\underbrace{q}_{\in\{1,2,\ldots,i-1\}}\right) \in \sigma\left(\{1, 2, \ldots, i-1\}\right)$.

Let us now forget that we fixed p. We thus have shown that every $p \in \left\{j \in \{1, 2, \ldots, n\} \ \mid \ a_j \leq y\right\}$ satisfies $p \in \sigma\left(\{1, 2, \ldots, i-1\}\right)$. In other words,

$$\left\{j \in \{1, 2, \ldots, n\} \ \mid \ a_j \leq y\right\} \subseteq \sigma\left(\{1, 2, \ldots, i-1\}\right).$$

Hence,

$$\begin{aligned} \left|\left\{j \in \{1, 2, \ldots, n\} \ \mid \ a_j \leq y\right\}\right| &\leq \left|\sigma\left(\{1, 2, \ldots, i-1\}\right)\right| = \left|\{1, 2, \ldots, i-1\}\right| \\ &\qquad \left(\text{since the map } \sigma \text{ is injective}\right) \\ &= i - 1 < i, \end{aligned}$$

qed.

- There is **at least one** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$ [497].
- There is **at most one** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$ [498].

Combining these two statements, we conclude that there is a **unique** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. This proves Proposition 6.40 **(c)**. □

Proof of Lemma 6.41. We have

$$\mathbf{E} = \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct} \right\}$$

and

$$\mathbf{I} = \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid k_1 < k_2 < \cdots < k_n \right\}.$$

Clearly, every element of $\mathbf{I} \times S_n$ can be written in the form $((g_1, g_2, \ldots, g_n), \sigma)$ for some $(g_1, g_2, \ldots, g_n) \in \mathbf{I}$ and $\sigma \in S_n$.

---

[497] *Proof.* Proposition 6.40 **(a)** shows that there exists a permutation $\sigma \in S_n$ such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. Consider this σ.

Let $k \in \{1, 2, \ldots, n-1\}$. Then, $a_{\sigma(k)} \leq a_{\sigma(k+1)}$ (since $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$). But on the other hand, $k \neq k+1$. Hence, $\sigma(k) \neq \sigma(k+1)$ (since the map σ is injective). Hence, $a_{\sigma(k)} \neq a_{\sigma(k+1)}$ (by (1082), applied to $u = \sigma(k)$ and $v = \sigma(k+1)$). Combined with $a_{\sigma(k)} \leq a_{\sigma(k+1)}$, this yields $a_{\sigma(k)} < a_{\sigma(k+1)}$.

Let us now forget that we fixed k. We thus have shown that $a_{\sigma(k)} < a_{\sigma(k+1)}$ for every $k \in \{1, 2, \ldots, n-1\}$. In other words, $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$.

Let us now forget that we defined σ. We thus have constructed a permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. Hence, there is **at least one** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. Qed.

[498] *Proof.* Let σ_1 and σ_2 be two permutations $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. We shall show that $\sigma_1 = \sigma_2$.

Fix $i \in \{1, 2, \ldots, n\}$.

We know that σ_1 is a permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. In other words, σ_1 is a permutation in S_n such that $a_{\sigma_1(1)} < a_{\sigma_1(2)} < \cdots < a_{\sigma_1(n)}$. Thus, (1079) (applied to $\sigma = \sigma_1$) yields

$$a_{\sigma_1(i)} = \min \left\{ x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x \right\}. \tag{1083}$$

The same argument (applied to σ_2 instead of σ_1) yields

$$a_{\sigma_2(i)} = \min \left\{ x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x \right\}.$$

Comparing this with (1083), we obtain $a_{\sigma_1(i)} = a_{\sigma_2(i)}$.

Now, if we had $\sigma_1(i) \neq \sigma_2(i)$, then we would have $a_{\sigma_1(i)} \neq a_{\sigma_2(i)}$ (by (1082), applied to $u = \sigma_1(i)$ and $v = \sigma_2(i)$), which would contradict $a_{\sigma_1(i)} = a_{\sigma_2(i)}$. Thus, we cannot have $\sigma_1(i) \neq \sigma_2(i)$. Hence, we must have $\sigma_1(i) = \sigma_2(i)$.

Let us now forget that we fixed i. We thus have shown that $\sigma_1(i) = \sigma_2(i)$ for every $i \in \{1, 2, \ldots, n\}$. In other words, $\sigma_1 = \sigma_2$.

Let us now forget that we fixed σ_1 and σ_2. We thus have shown that $\sigma_1 = \sigma_2$ whenever σ_1 and σ_2 are two permutations $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. In other words, there is **at most one** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. Qed.

For every $((g_1, g_2, \ldots, g_n), \sigma) \in \mathbf{I} \times S_n$, we have $\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right) \in \mathbf{E}$ [499]. Hence, we can define a map $\Phi : \mathbf{I} \times S_n \to \mathbf{E}$ by setting

$$\left(\begin{array}{c} \Phi((g_1, g_2, \ldots, g_n), \sigma) = \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right) \\ \text{for every } ((g_1, g_2, \ldots, g_n), \sigma) \in \mathbf{I} \times S_n \end{array} \right) \tag{1085}$$

(since every element of $\mathbf{I} \times S_n$ can be written in the form $((g_1, g_2, \ldots, g_n), \sigma)$ for some $(g_1, g_2, \ldots, g_n) \in \mathbf{I}$ and $\sigma \in S_n$). Consider this map Φ.

The map Φ is the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

(since Φ is defined by (1085)). In particular, the latter map is well-defined.

---

[499] *Proof.* Let $((g_1, g_2, \ldots, g_n), \sigma) \in \mathbf{I} \times S_n$. Thus, $(g_1, g_2, \ldots, g_n) \in \mathbf{I}$ and $\sigma \in S_n$.

We have $\sigma \in S_n$. Thus, σ is a permutation, and thus a bijective map, hence an injective map.

We have

$$(g_1, g_2, \ldots, g_n) \in \mathbf{I} = \left\{(k_1, k_2, \ldots, k_n) \in [m]^n \mid k_1 < k_2 < \cdots < k_n\right\}.$$

In other words, $(g_1, g_2, \ldots, g_n)$ is an element $(k_1, k_2, \ldots, k_n) \in [m]^n$ satisfying $k_1 < k_2 < \cdots < k_n$. In other words, $(g_1, g_2, \ldots, g_n)$ is an element of $[m]^n$ and satisfies $g_1 < g_2 < \cdots < g_n$. Hence, the integers $g_1, g_2, \ldots, g_n$ are distinct (since $g_1 < g_2 < \cdots < g_n$). In other words, any two distinct elements u and v of $[n]$ satisfy

$$g_u \neq g_v. \tag{1084}$$

Now, let u and v be two distinct elements of $[n]$. Thus, $u \neq v$ (since u and v are distinct), so that $\sigma(u) \neq \sigma(v)$ (since σ is injective). Hence, $g_{\sigma(u)} \neq g_{\sigma(v)}$ (by (1084), applied to $\sigma(u)$ and $\sigma(v)$ instead of u and v).

Let us now forget that we fixed u and v. We thus have shown that any two distinct elements u and v of $[n]$ satisfy $g_{\sigma(u)} \neq g_{\sigma(v)}$. In other words, the integers $g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}$ are distinct. Hence, $\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$ is an element $(k_1, k_2, \ldots, k_n) \in [m]^n$ such that the integers $k_1, k_2, \ldots, k_n$ are distinct. In other words,

$$\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right) \in \left\{(k_1, k_2, \ldots, k_n) \in [m]^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct}\right\} = \mathbf{E},$$

qed.

The map Φ is injective[500] and surjective[501]. Hence, the map Φ is bijective. In

[500] *Proof.* Let α and β be two elements of $\mathbf{I} \times S_n$ such that $\Phi(\alpha) = \Phi(\beta)$. We shall show that $\alpha = \beta$.

Let γ be the element $\Phi(\alpha) = \Phi(\beta)$ of $\mathbf{E}$. Thus, $\gamma = \Phi(\alpha) = \Phi(\beta)$.

Write $\alpha \in \mathbf{I} \times S_n$ in the form $\alpha = ((g_1, g_2, \ldots, g_n), \pi)$ for some $(g_1, g_2, \ldots, g_n) \in \mathbf{I}$ and $\pi \in S_n$. Write $\beta \in \mathbf{I} \times S_n$ in the form $\beta = ((h_1, h_2, \ldots, h_n), \tau)$ for some $(h_1, h_2, \ldots, h_n) \in \mathbf{I}$ and $\tau \in S_n$.

We have

$$(g_1, g_2, \ldots, g_n) \in \mathbf{I} = \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid k_1 < k_2 < \cdots < k_n \right\}.$$

In other words, $(g_1, g_2, \ldots, g_n)$ is an element $(k_1, k_2, \ldots, k_n)$ of $[m]^n$ satisfying $k_1 < k_2 < \cdots < k_n$. In other words, $(g_1, g_2, \ldots, g_n)$ is an element of $[m]^n$ and satisfies $g_1 < g_2 < \cdots < g_n$.

We have

$$(h_1, h_2, \ldots, h_n) \in \mathbf{I} = \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid k_1 < k_2 < \cdots < k_n \right\}.$$

In other words, $(h_1, h_2, \ldots, h_n)$ is an element $(k_1, k_2, \ldots, k_n)$ of $[m]^n$ satisfying $k_1 < k_2 < \cdots < k_n$. In other words, $(h_1, h_2, \ldots, h_n)$ is an element of $[m]^n$ and satisfies $h_1 < h_2 < \cdots < h_n$.

We have $\gamma \in \mathbf{E} = \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct} \right\}$. In other words, we can write γ in the form $\gamma = (k_1, k_2, \ldots, k_n)$ for some $(k_1, k_2, \ldots, k_n) \in [m]^n$ such that the integers $k_1, k_2, \ldots, k_n$ are distinct. Consider this $(k_1, k_2, \ldots, k_n)$.

Applying the map Φ to both sides of the equality $\alpha = ((g_1, g_2, \ldots, g_n), \pi)$, we obtain

$$\Phi(\alpha) = \Phi((g_1, g_2, \ldots, g_n), \pi) = \left(g_{\pi(1)}, g_{\pi(2)}, \ldots, g_{\pi(n)} \right) \qquad (\text{by the definition of } \Phi).$$

Thus, $\left(g_{\pi(1)}, g_{\pi(2)}, \ldots, g_{\pi(n)} \right) = \Phi(\alpha) = \gamma = (k_1, k_2, \ldots, k_n)$. In other words, every $i \in \{1, 2, \ldots, n\}$ satisfies

$$g_{\pi(i)} = k_i. \tag{1086}$$

Thus, every $j \in \{1, 2, \ldots, n\}$ satisfies

$$\begin{aligned} g_j &= g_{\pi(\pi^{-1}(j))} && \left(\text{since } j = \pi\left(\pi^{-1}(j)\right)\right) \\ &= k_{\pi^{-1}(j)} && \left(\text{by (1086), applied to } i = \pi^{-1}(j)\right). \end{aligned} \tag{1087}$$

We have $g_1 < g_2 < \cdots < g_n$. This rewrites as

$$k_{\pi^{-1}(1)} < k_{\pi^{-1}(2)} < \cdots < k_{\pi^{-1}(n)}$$

(because every $j \in \{1, 2, \ldots, n\}$ satisfies $g_j = k_{\pi^{-1}(j)}$).

Applying the map Φ to both sides of the equality $\beta = ((h_1, h_2, \ldots, h_n), \tau)$, we obtain

$$\Phi(\beta) = \Phi((h_1, h_2, \ldots, h_n), \tau) = \left(h_{\tau(1)}, h_{\tau(2)}, \ldots, h_{\tau(n)} \right) \qquad (\text{by the definition of } \Phi).$$

Thus, $\left(h_{\tau(1)}, h_{\tau(2)}, \ldots, h_{\tau(n)} \right) = \Phi(\beta) = \gamma = (k_1, k_2, \ldots, k_n)$. In other words, every $i \in \{1, 2, \ldots, n\}$ satisfies

$$h_{\tau(i)} = k_i. \tag{1088}$$

Thus, every $j \in \{1, 2, \ldots, n\}$ satisfies

$$\begin{aligned} h_j &= h_{\tau(\tau^{-1}(j))} && \left(\text{since } j = \tau\left(\tau^{-1}(j)\right)\right) \\ &= k_{\tau^{-1}(j)} && \left(\text{by (1088), applied to } i = \tau^{-1}(j)\right). \end{aligned} \tag{1089}$$

We have $h_1 < h_2 < \cdots < h_n$. This rewrites as

$$k_{\tau^{-1}(1)} < k_{\tau^{-1}(2)} < \cdots < k_{\tau^{-1}(n)}$$

(because every $j \in \{1, 2, \ldots, n\}$ satisfies $h_j = k_{\tau^{-1}(j)}$).

Proposition 6.40 **(c)** (applied to $a_i = k_i$) yields that there is a **unique** permutation $\sigma \in S_n$ such that $k_{\sigma(1)} < k_{\sigma(2)} < \cdots < k_{\sigma(n)}$ (since the integers $k_1, k_2, \ldots, k_n$ are distinct). In particular, there exists **at most one** such permutation. In other words, if σ_1 and σ_2 are two permutations $\sigma \in S_n$ satisfying $k_{\sigma(1)} < k_{\sigma(2)} < \cdots < k_{\sigma(n)}$, then

$$\sigma_1 = \sigma_2. \tag{1090}$$

Now, π^{-1} is a permutation $\sigma \in S_n$ satisfying $k_{\sigma(1)} < k_{\sigma(2)} < \cdots < k_{\sigma(n)}$ (since $k_{\pi^{-1}(1)} < k_{\pi^{-1}(2)} < \cdots < k_{\pi^{-1}(n)}$). Also, τ^{-1} is a permutation $\sigma \in S_n$ satisfying $k_{\sigma(1)} < k_{\sigma(2)} < \cdots < k_{\sigma(n)}$ (since $k_{\tau^{-1}(1)} < k_{\tau^{-1}(2)} < \cdots < k_{\tau^{-1}(n)}$). Hence, we can apply (1090) to $\sigma_1 = \pi^{-1}$ and $\sigma_2 = \tau^{-1}$. As a result, we obtain $\pi^{-1} = \tau^{-1}$. Hence, $\pi = \tau$. Now, every $j \in \{1, 2, \ldots, n\}$ satisfies

$$\begin{aligned} g_j &= k_{\pi^{-1}(j)} \qquad (\text{by (1087)}) \\ &= k_{\tau^{-1}(j)} \qquad (\text{since } \pi = \tau) \\ &= h_j \qquad (\text{by (1089)}). \end{aligned}$$

In other words, $(g_1, g_2, \ldots, g_n) = (h_1, h_2, \ldots, h_n)$. Thus,

$$\alpha = \left(\underbrace{(g_1, g_2, \ldots, g_n)}_{=(h_1, h_2, \ldots, h_n)}, \underbrace{\pi}_{=\tau} \right) = ((h_1, h_2, \ldots, h_n), \tau) = \beta.$$

Now, let us forget that we fixed α and β. We thus have shown that if α and β are two elements of $\mathbf{I} \times S_n$ such that $\Phi(\alpha) = \Phi(\beta)$, then $\alpha = \beta$. In other words, the map Φ is injective, qed.

[501] *Proof.* Let $\gamma \in \mathbf{E}$. We shall prove that $\gamma \in \Phi(\mathbf{I} \times S_n)$.

We have $\gamma \in \mathbf{E} = \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct} \right\}$. In other words, we can write γ in the form $\gamma = (k_1, k_2, \ldots, k_n)$ for some $(k_1, k_2, \ldots, k_n) \in [m]^n$ such that the integers $k_1, k_2, \ldots, k_n$ are distinct. Let us denote this $(k_1, k_2, \ldots, k_n)$ by $(a_1, a_2, \ldots, a_n)$. Thus, $(a_1, a_2, \ldots, a_n)$ is an element of $[m]^n$ such that the integers $a_1, a_2, \ldots, a_n$ are distinct, and we have $\gamma = (a_1, a_2, \ldots, a_n)$.

Proposition 6.40 **(c)** yields that there is a **unique** permutation $\sigma \in S_n$ such that $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$ (since the integers $a_1, a_2, \ldots, a_n$ are distinct). In particular, there exists **at least one** such permutation. Consider such a permutation σ. Thus, $\sigma \in S_n$ and $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$.

Now, $\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right)$ is an element of $[m]^n$ satisfying $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. In other words, $\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right)$ is an element $(k_1, k_2, \ldots, k_n)$ of $[m]^n$ satisfying $k_1 < k_2 < \cdots < k_n$. In other words,

$$\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right) \in \left\{ (k_1, k_2, \ldots, k_n) \in [m]^n \mid k_1 < k_2 < \cdots < k_n \right\} = \mathbf{I}.$$

Hence, $\Phi\left(\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right), \sigma^{-1}\right)$ is well-defined. The definition of other words, the map Φ is a bijection. In other words, the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

is a bijection (because the map Φ is the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

). This concludes the proof of Lemma 6.41. □

7.81. Solution to Exercise 6.14

Before we start solving Exercise 6.14, let us isolate a useful fact that was proven in our above proof of Proposition 6.40 **(b)**:

Lemma 7.162. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n integers. Let $\sigma \in S_n$ be such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. Let $i \in \{1, 2, \ldots, n\}$. Then,

$$a_{\sigma(i)} = \min \left\{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\right\} \qquad (1091)$$

(and, in particular, the right hand side of (1091) is well-defined).

Proof of Lemma 7.162. The equality (1091) is precisely the equality (1079) that was proven during our above proof of Proposition 6.40 **(b)**. Hence, we do not need to prove it again. Thus, Lemma 7.162 is proven. □

$\Phi\left(\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right), \sigma^{-1}\right)$ yields

$$\begin{aligned}
&\Phi\left(\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right), \sigma^{-1}\right) \\
&= \left(a_{\sigma(\sigma^{-1}(1))}, a_{\sigma(\sigma^{-1}(2))}, \ldots, a_{\sigma(\sigma^{-1}(n))}\right) = (a_1, a_2, \ldots, a_n) \\
&\qquad \left(\text{since } a_{\sigma(\sigma^{-1}(i))} = a_i \text{ for every } i \in \{1, 2, \ldots, n\}\right) \\
&= \gamma.
\end{aligned}$$

Thus, $\gamma = \Phi\left(\underbrace{\left(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}\right)}_{\in \mathbf{I}}, \underbrace{\sigma^{-1}}_{\in S_n}\right) \in \Phi(\mathbf{I} \times S_n)$.

Now, let us forget that we fixed γ. We thus have shown that $\gamma \in \Phi(\mathbf{I} \times S_n)$ for every $\gamma \in \mathbf{E}$. In other words, $\mathbf{E} \subseteq \Phi(\mathbf{I} \times S_n)$. In other words, the map Φ is surjective. Qed.

First solution to Exercise 6.14. Let $i \in \{1, 2, \ldots, m\}$. We must prove that $a_{\sigma(i)} \leq b_{\tau(i)}$.

We have $n \geq m$, so that $m \leq n$. Now, $i \in \{1, 2, \ldots, m\} \subseteq \{1, 2, \ldots, n\}$ (since $m \leq n$). Hence, $\sigma(i)$ is well-defined (since σ is a map $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ (since $\sigma \in S_n$)).

Define a subset X of $\mathbb{Z}$ by

$$X = \{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\}. \tag{1092}$$

Define a subset Y of $\mathbb{Z}$ by

$$Y = \{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, m\} \text{ satisfy } b_j \leq x\}. \tag{1093}$$

Lemma 7.162 yields that

$$a_{\sigma(i)} = \min \{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\} \tag{1094}$$

(and, in particular, the right hand side of (1094) is well-defined). Thus,

$$a_{\sigma(i)} = \min \underbrace{\{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, n\} \text{ satisfy } a_j \leq x\}}_{\substack{=X \\ \text{(by (1092))}}} = \min X. \tag{1095}$$

But $\tau \in S_m$ satisfies $b_{\tau(1)} \leq b_{\tau(2)} \leq \cdots \leq b_{\tau(m)}$, and we have $i \in \{1, 2, \ldots, m\}$. Thus, Lemma 7.162 (applied to m, b_k and τ instead of n, a_k and σ) yields that

$$b_{\tau(i)} = \min \{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, m\} \text{ satisfy } b_j \leq x\} \tag{1096}$$

(and, in particular, the right hand side of (1096) is well-defined). Thus,

$$b_{\tau(i)} = \min \underbrace{\{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, m\} \text{ satisfy } b_j \leq x\}}_{\substack{=Y \\ \text{(by (1093))}}} = \min Y. \tag{1097}$$

But clearly, $\min Y \in Y$. Hence,

$$b_{\tau(i)} = \min Y \in Y = \{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1, 2, \ldots, m\} \text{ satisfy } b_j \leq x\}.$$

In other words, $b_{\tau(i)}$ is an element of $\mathbb{Z}$ such that at least i elements $j \in \{1, 2, \ldots, m\}$ satisfy $b_j \leq b_{\tau(i)}$.

We now know that at least i elements $j \in \{1, 2, \ldots, m\}$ satisfy $b_j \leq b_{\tau(i)}$. Each of these i elements j must also be an element of $\{1, 2, \ldots, n\}$ (since $j \in \{1, 2, \ldots, m\} \subseteq \{1, 2, \ldots, n\}$) which satisfies $a_j \leq b_{\tau(i)}$ (because (383) (applied to j instead of i) shows that $a_j \leq b_j \leq b_{\tau(i)}$). Thus, at least i elements $j \in \{1, 2, \ldots, n\}$ satisfy $a_j \leq b_{\tau(i)}$.

Now, we know that $b_{\tau(i)}$ is an element of $\mathbb{Z}$ such that at least i elements $j \in \{1,2,\ldots,n\}$ satisfy $a_j \leq b_{\tau(i)}$. In other words, $b_{\tau(i)}$ is an element x of $\mathbb{Z}$ such that at least i elements $j \in \{1,2,\ldots,n\}$ satisfy $a_j \leq x$. In other words,

$$b_{\tau(i)} \in \left\{x \in \mathbb{Z} \mid \text{at least } i \text{ elements } j \in \{1,2,\ldots,n\} \text{ satisfy } a_j \leq x\right\}.$$

In light of (1092), this rewrites as $b_{\tau(i)} \in X$.

But every $s \in X$ satisfies $\min X \leq s$ (since the minimum of a set is $\leq$ to each element of this set). Applying this to $s = b_{\tau(i)}$, we obtain $\min X \leq b_{\tau(i)}$ (since $b_{\tau(i)} \in X$). Thus, $b_{\tau(i)} \geq \min X = a_{\sigma(i)}$ (by (1095)). This solves Exercise 6.14. $\square$

In order to give a second solution to Exercise 6.14, let us prove a lemma:

Lemma 7.163. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that $n \geq m$. Let $a_1, a_2, \ldots, a_n$ be n integers. Let $b_1, b_2, \ldots, b_m$ be m integers. Assume that

$$a_i \leq b_i \qquad \text{for every } i \in \{1,2,\ldots,m\}. \tag{1098}$$

Let $\sigma \in S_n$ and $\tau \in S_m$. Let $i \in \{1,2,\ldots,m\}$. Assume that

$$a_{\sigma(i)} \leq a_{\sigma(v)} \qquad \text{for every } v \in \{i, i+1, \ldots, n\}. \tag{1099}$$

Furthermore, assume that

$$b_{\tau(u)} \leq b_{\tau(i)} \qquad \text{for every } u \in \{1,2,\ldots,i\}. \tag{1100}$$

Then, $a_{\sigma(i)} \leq b_{\tau(i)}$.

Proof of Lemma 7.163. We have $i \in \{1,2,\ldots,m\} \subseteq \{1,2,\ldots,n\}$ (since $m \leq n$).

We have $\sigma \in S_n$. Thus, σ is a permutation of $\{1,2,\ldots,n\}$, hence an injective map. Thus, the n integers $\sigma(1), \sigma(2), \ldots, \sigma(n)$ are distinct. In particular, the $n-i+1$ integers $\sigma(i), \sigma(i+1), \ldots, \sigma(n)$ are distinct.

Also, $\tau \in S_m$. Therefore, τ is a permutation of $\{1,2,\ldots,m\}$, hence an injective map. Therefore, the m integers $\tau(1), \tau(2), \ldots, \tau(m)$ are distinct. In particular, the i integers $\tau(1), \tau(2), \ldots, \tau(i)$ are distinct.

If A and B are two subsets of $\{1,2,\ldots,n\}$ satisfying $|A| + |B| > n$, then

$$A \cap B \neq \varnothing \tag{1101}$$

[502].

[502] *Proof of (1101):* Let A and B be two subsets of $\{1,2,\ldots,n\}$ satisfying $|A| + |B| > n$. We must prove that $A \cap B \neq \varnothing$.

Indeed, assume the contrary. Thus, $A \cap B = \varnothing$. We know that A and B are subsets of $\{1,2,\ldots,n\}$. Hence, $A \cup B$ is a subset of $\{1,2,\ldots,n\}$. Thus, $|A \cup B| \leq |\{1,2,\ldots,n\}| = n$. But the sets A and B are disjoint (since $A \cap B = \varnothing$), and thus we have $|A \cup B| = |A| + |B|$ (since the size of the union of two disjoint sets is the sum of their sizes). Thus, $|A| + |B| = |A \cup B| \leq n$. This contradicts $|A| + |B| > n$. This contradiction proves that our assumption was wrong. Hence, $A \cap B \neq \varnothing$ is proven. This proves (1101).

Now, let $A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(n)\}$ and $B = \{\tau(1), \tau(2), \ldots, \tau(i)\}$. Clearly,

$$A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(n)\} \subseteq \{1, 2, \ldots, n\} \qquad (1102)$$

(since σ is a map $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$) and $B = \{\tau(1), \tau(2), \ldots, \tau(i)\} \subseteq \{1, 2, \ldots, m\}$ (since τ is a map $\{1, 2, \ldots, m\} \to \{1, 2, \ldots, m\}$). Thus,

$$B \subseteq \{1, 2, \ldots, m\} \subseteq \{1, 2, \ldots, n\}. \qquad (1103)$$

Now, from (1102) and (1103), we see that A and B are two subsets of $\{1, 2, \ldots, n\}$. These two subsets satisfy

$$\begin{aligned} & \left| \underbrace{A}_{=\{\sigma(i),\sigma(i+1),\ldots,\sigma(n)\}} \right| + \left| \underbrace{B}_{=\{\tau(1),\tau(2),\ldots,\tau(i)\}} \right| \\ &= \underbrace{\left|\{\sigma(i), \sigma(i+1), \ldots, \sigma(n)\}\right|}_{\substack{=n-i+1 \\ \text{(since the } n-i+1 \text{ integers} \\ \sigma(i),\sigma(i+1),\ldots,\sigma(n) \text{ are distinct)}}} + \underbrace{\left|\{\tau(1), \tau(2), \ldots, \tau(i)\}\right|}_{\substack{=i \\ \text{(since the } i \text{ integers} \\ \tau(1),\tau(2),\ldots,\tau(i) \text{ are distinct)}}} \\ &= n - i + 1 + i = n + 1 > n. \end{aligned}$$

Hence, (1101) shows that $A \cap B \neq \varnothing$. In other words, there exists a $g \in A \cap B$. Consider this g.

We have $g \in A \cap B \subseteq B = \{\tau(1), \tau(2), \ldots, \tau(i)\} \subseteq \{1, 2, \ldots, m\}$. Hence, b_g is well-defined. Also, $g \in \{\tau(1), \tau(2), \ldots, \tau(i)\}$. In other words, there exists a $u \in \{1, 2, \ldots, i\}$ satisfying $g = \tau(u)$. Consider this u. From (1100), we obtain $b_{\tau(u)} \leq b_{\tau(i)}$. But $g = \tau(u)$ shows that $b_g = b_{\tau(u)} \leq b_{\tau(i)}$.

On the other hand, $g \in A \cap B \subseteq A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(n)\} \subseteq \{1, 2, \ldots, n\}$. Hence, a_g is well-defined. Also, $g \in \{\sigma(i), \sigma(i+1), \ldots, \sigma(n)\}$. In other words, there exists a $v \in \{i, i+1, \ldots, n\}$ such that $g = \sigma(v)$. Consider this v. From (1099), we obtain $a_{\sigma(i)} \leq a_{\sigma(v)} = a_g$ (since $\sigma(v) = g$).

But (1098) (applied to g instead of i) shows that $a_g \leq b_g$ (since $g \in \{1, 2, \ldots, m\}$). Hence, $a_{\sigma(i)} \leq a_g \leq b_g \leq b_{\tau(i)}$. This proves Lemma 7.163. □

Second solution to Exercise 6.14. Let $i \in \{1, 2, \ldots, m\}$. We must prove that $a_{\sigma(i)} \leq b_{\tau(i)}$.

We have $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$. In other words, for every $u \in \{1, 2, \ldots, n\}$ and $v \in \{1, 2, \ldots, n\}$ satisfying $u \leq v$, we have

$$a_{\sigma(u)} \leq a_{\sigma(v)}. \qquad (1104)$$

Also, we have $b_{\tau(1)} \leq b_{\tau(2)} \leq \cdots \leq b_{\tau(m)}$. In other words, for every $u \in \{1, 2, \ldots, m\}$ and $v \in \{1, 2, \ldots, m\}$ satisfying $u \leq v$, we have

$$b_{\tau(u)} \leq b_{\tau(v)}. \qquad (1105)$$

Now,

$$a_{\sigma(i)} \leq a_{\sigma(v)} \qquad \text{for every } v \in \{i, i+1, \ldots, n\}$$

[503]. Furthermore,

$$b_{\tau(u)} \leq b_{\tau(i)} \qquad \text{for every } u \in \{1, 2, \ldots, i\}$$

[504]. Hence, Lemma 7.163 shows that $a_{\sigma(i)} \leq b_{\tau(i)}$. This solves Exercise 6.14 again. $\square$

7.82. Solution to Exercise 6.15

Definition 6.50 shall be used throughout this section.

Proof of Lemma 6.52. **(a)** Let k be a negative integer. Then,

$$h_k(x_1, x_2, \ldots, x_n) = \sum_{\substack{(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n; \\ a_1 + a_2 + \cdots + a_n = k}} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \tag{1106}$$

(by the definition of $h_k(x_1, x_2, \ldots, x_n)$).

But $k < 0$. Hence, there exists no $(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n$ satisfying $a_1 + a_2 + \cdots + a_n = k$ (because every $(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n$ satisfies $a_1 + a_2 + \cdots + a_n \geq 0$, whereas $k < 0$). Therefore, the sum on the right hand side of (1106) is an empty sum and therefore equals 0. Therefore, (1106) simplifies to $h_k(x_1, x_2, \ldots, x_n) = 0$. This proves Lemma 6.52 **(a)**.

(b) The definition of $h_0(x_1, x_2, \ldots, x_n)$ yields

$$h_0(x_1, x_2, \ldots, x_n) = \sum_{\substack{(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n; \\ a_1 + a_2 + \cdots + a_n = 0}} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}. \tag{1107}$$

But the only $(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n$ satisfying $a_1 + a_2 + \cdots + a_n = 0$ is $\left(\underbrace{0, 0, \ldots, 0}_{n \text{ zeroes}}\right)$ (since a sum of nonnegative integers can only be 0 if all addends are 0). Therefore, the sum on the right hand side of (1107) has exactly one addend – namely, the addend for $(a_1, a_2, \ldots, a_n) = \left(\underbrace{0, 0, \ldots, 0}_{n \text{ zeroes}}\right)$. Thus, (1107) simplifies to $h_0(x_1, x_2, \ldots, x_n) = \underbrace{x_1^0}_{=1} \underbrace{x_2^0}_{=1} \cdots \underbrace{x_n^0}_{=1} = 1$. This proves Lemma 6.52 **(b)**. $\square$

[503]*Proof.* Let $v \in \{i, i+1, \ldots, n\}$. Then, $v \geq i$, so that $i \leq v$. Also, $i \in \{1, 2, \ldots, m\} \subseteq \{1, 2, \ldots, n\}$ (since $m \leq n$ (since $n \geq m$)) and $v \in \{i, i+1, \ldots, n\} \subseteq \{1, 2, \ldots, n\}$ (since $i \geq 1$ (since $i \in \{1, 2, \ldots, m\}$)). Hence, (1104) (applied to $u = i$) yields $a_{\sigma(i)} \leq a_{\sigma(v)}$. Qed.

[504]*Proof.* Let $u \in \{1, 2, \ldots, i\}$. Then, $u \leq i$. Also, $u \in \{1, 2, \ldots, i\} \subseteq \{1, 2, \ldots, m\}$ (since $i \leq m$ (since $i \in \{1, 2, \ldots, m\}$)) and $i \in \{1, 2, \ldots, m\}$. Hence, (1105) (applied to $v = i$) yields $b_{\tau(u)} \leq b_{\tau(i)}$. Qed.

Proof of Lemma 6.53. The definition of $h_q(x_1, x_2, \ldots, x_k)$ yields

$$h_q(x_1, x_2, \ldots, x_k) = \sum_{\substack{(a_1,a_2,\ldots,a_k)\in\mathbb{N}^k;\\ a_1+a_2+\cdots+a_k=q}} x_1^{a_1} x_2^{a_2} \cdots x_k^{a_k} = \sum_{\substack{(s_1,s_2,\ldots,s_k)\in\mathbb{N}^k;\\ s_1+s_2+\cdots+s_k=q}} x_1^{s_1} x_2^{s_2} \cdots x_k^{s_k} \qquad (1108)$$

(here, we have renamed the summation index $(a_1, a_2, \ldots, a_k)$ as $(s_1, s_2, \ldots, s_k)$).

Every $r \in \mathbb{Z}$ satisfies

$$h_{q-r}(x_1, x_2, \ldots, x_{k-1}) = \sum_{\substack{(a_1,a_2,\ldots,a_{k-1})\in\mathbb{N}^{k-1};\\ a_1+a_2+\cdots+a_{k-1}=q-r}} x_1^{a_1} x_2^{a_2} \cdots x_{k-1}^{a_{k-1}}$$
$$\left(\text{by the definition of } h_{q-r}(x_1, x_2, \ldots, x_{k-1})\right)$$
$$= \sum_{\substack{(s_1,s_2,\ldots,s_{k-1})\in\mathbb{N}^{k-1};\\ s_1+s_2+\cdots+s_{k-1}=q-r}} x_1^{s_1} x_2^{s_2} \cdots x_{k-1}^{s_{k-1}} \qquad (1109)$$

(here, we have renamed the summation index $(a_1, a_2, \ldots, a_{k-1})$ as $(s_1, s_2, \ldots, s_{k-1})$). Moreover, every $r \in \mathbb{Z}$ satisfying $r > q$ satisfies

$$h_{q-r}(x_1, x_2, \ldots, x_{k-1}) = 0 \qquad (1110)$$

[505].

But Corollary 7.19 (applied to $M = k$ and $Z = \mathbb{N}$) shows that the map

$$\mathbb{N}^k \to \mathbb{N}^{k-1} \times \mathbb{N},$$
$$(s_1, s_2, \ldots, s_k) \mapsto ((s_1, s_2, \ldots, s_{k-1}), s_k)$$

is a bijection.

[505] *Proof of (1110):* Let $r \in \mathbb{Z}$ be such that $r > q$. The integer $q - r$ is negative (since $r > q$). Hence, Lemma 6.52 **(a)** (applied to $k-1$ and $q-r$ instead of n and k) yields $h_{q-r}(x_1, x_2, \ldots, x_{k-1}) = 0$. This proves (1110).

Now, (1108) becomes

$$\begin{aligned}
h_q(x_1,x_2,\ldots,x_k) &= \underbrace{\sum_{\substack{(s_1,s_2,\ldots,s_k)\in\mathbb{N}^k;\\ s_1+s_2+\cdots+s_k=q}}}_{\substack{=\sum_{\substack{(s_1,s_2,\ldots,s_k)\in\mathbb{N}^k;\\ s_1+s_2+\cdots+s_{k-1}=q-s_k}}\\ \text{(because for any } (s_1,s_2,\ldots,s_k)\in\mathbb{N}^k\text{,}\\ \text{the condition } s_1+s_2+\cdots+s_k=q\\ \text{is equivalent to } s_1+s_2+\cdots+s_{k-1}=q-s_k)}} \underbrace{x_1^{s_1}x_2^{s_2}\cdots x_k^{s_k}}_{=\left(x_1^{s_1}x_2^{s_2}\cdots x_{k-1}^{s_{k-1}}\right)x_k^{s_k}} \\
&= \sum_{\substack{(s_1,s_2,\ldots,s_k)\in\mathbb{N}^k;\\ s_1+s_2+\cdots+s_{k-1}=q-s_k}} \left(x_1^{s_1}x_2^{s_2}\cdots x_{k-1}^{s_{k-1}}\right)x_k^{s_k} \\
&= \underbrace{\sum_{\substack{((s_1,s_2,\ldots,s_{k-1}),r)\in\mathbb{N}^{k-1}\times\mathbb{N};\\ s_1+s_2+\cdots+s_{k-1}=q-r}}}_{=\sum_{r\in\mathbb{N}}\sum_{\substack{(s_1,s_2,\ldots,s_{k-1})\in\mathbb{N}^{k-1};\\ s_1+s_2+\cdots+s_{k-1}=q-r}}} \left(x_1^{s_1}x_2^{s_2}\cdots x_{k-1}^{s_{k-1}}\right)x_k^{r} \\
&\qquad \left(\begin{array}{c}\text{here, we have substituted } ((s_1,s_2,\ldots,s_{k-1}),r) \text{ for}\\ ((s_1,s_2,\ldots,s_{k-1}),s_k)\text{, since the map}\\ \mathbb{N}^k\to\mathbb{N}^{k-1}\times\mathbb{N},\ (s_1,s_2,\ldots,s_k)\mapsto((s_1,s_2,\ldots,s_{k-1}),s_k)\\ \text{is a bijection}\end{array}\right) \\
&= \sum_{r\in\mathbb{N}}\sum_{\substack{(s_1,s_2,\ldots,s_{k-1})\in\mathbb{N}^{k-1};\\ s_1+s_2+\cdots+s_{k-1}=q-r}} \left(x_1^{s_1}x_2^{s_2}\cdots x_{k-1}^{s_{k-1}}\right)x_k^{r} \\
&= \sum_{r\in\mathbb{N}} x_k^r \underbrace{\sum_{\substack{(s_1,s_2,\ldots,s_{k-1})\in\mathbb{N}^{k-1};\\ s_1+s_2+\cdots+s_{k-1}=q-r}} x_1^{s_1}x_2^{s_2}\cdots x_{k-1}^{s_{k-1}}}_{\substack{=h_{q-r}(x_1,x_2,\ldots,x_{k-1})\\ \text{(by (1109))}}} \\
&= \sum_{r\in\mathbb{N}} x_k^r h_{q-r}(x_1,x_2,\ldots,x_{k-1}) \\
&= \underbrace{\sum_{\substack{r\in\mathbb{N};\\ r\leq q}}}_{=\sum_{r=0}^{q}} x_k^r h_{q-r}(x_1,x_2,\ldots,x_{k-1}) + \sum_{\substack{r\in\mathbb{N};\\ r>q}} x_k^r \underbrace{h_{q-r}(x_1,x_2,\ldots,x_{k-1})}_{\substack{=0\\ \text{(by (1110))}}} \\
&= \sum_{r=0}^{q} x_k^r h_{q-r}(x_1,x_2,\ldots,x_{k-1}) + \underbrace{\sum_{\substack{r\in\mathbb{N};\\ r>q}} x_k^r 0}_{=0} = \sum_{r=0}^{q} x_k^r h_{q-r}(x_1,x_2,\ldots,x_{k-1}).
\end{aligned}$$

This proves Lemma 6.53. □

Proof of Lemma 6.54. If $q < 0$, then Lemma 6.54 holds[506]. Hence, for the rest of this proof, we WLOG assume that we don't have $q < 0$. Thus, $q \geq 0$.

Lemma 6.53 (applied to $q-1$ instead of q) yields

$$\begin{aligned} h_{q-1}(x_1, x_2, \ldots, x_k) &= \sum_{r=0}^{q-1} x_k^r h_{(q-1)-r}(x_1, x_2, \ldots, x_{k-1}) \\ &= \sum_{r=1}^{q} x_k^{r-1} \underbrace{h_{(q-1)-(r-1)}(x_1, x_2, \ldots, x_{k-1})}_{\substack{= h_{q-r}(x_1, x_2, \ldots, x_{k-1}) \\ (\text{since } (q-1)-(r-1)=q-r)}} \\ &\qquad (\text{here, we have substituted } r-1 \text{ for } r \text{ in the sum}) \\ &= \sum_{r=1}^{q} x_k^{r-1} h_{q-r}(x_1, x_2, \ldots, x_{k-1}). \end{aligned} \tag{1111}$$

[506] *Proof.* Assume that $q < 0$. Thus, both $q-1$ and q are negative integers. Hence, Lemma 6.52 **(a)** (applied to $q-1$ and k instead of k and n) yields $h_{q-1}(x_1, x_2, \ldots, x_k) = 0$. Also, Lemma 6.52 **(a)** (applied to q and $k-1$ instead of k and n) yields $h_q(x_1, x_2, \ldots, x_{k-1}) = 0$.

Now, Lemma 6.52 **(a)** (applied to q and k instead of k and n) yields $h_q(x_1, x_2, \ldots, x_k) = 0$. Comparing this with

$$\underbrace{h_q(x_1, x_2, \ldots, x_{k-1})}_{=0} + x_k \underbrace{h_{q-1}(x_1, x_2, \ldots, x_k)}_{=0} = 0 + x_k 0 = 0,$$

we obtain

$$h_q(x_1, x_2, \ldots, x_k) = h_q(x_1, x_2, \ldots, x_{k-1}) + x_k h_{q-1}(x_1, x_2, \ldots, x_k).$$

Thus, Lemma 6.54 holds. Qed.

But Lemma 6.53 yields

$$\begin{aligned}
h_q(x_1, x_2, \ldots, x_k) &= \sum_{r=0}^{q} x_k^r h_{q-r}(x_1, x_2, \ldots, x_{k-1}) \\
&= \underbrace{x_k^0}_{=1} \underbrace{h_{q-0}(x_1, x_2, \ldots, x_{k-1})}_{=h_q(x_1, x_2, \ldots, x_{k-1})} + \sum_{r=1}^{q} \underbrace{x_k^r}_{=x_k x_k^{r-1}} h_{q-r}(x_1, x_2, \ldots, x_{k-1}) \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } r=0 \text{ from} \\ \text{the sum (since } q \geq 0\text{)}\end{array}\right) \\
&= h_q(x_1, x_2, \ldots, x_{k-1}) + \underbrace{\sum_{r=1}^{q} x_k x_k^{r-1} h_{q-r}(x_1, x_2, \ldots, x_{k-1})}_{=x_k \sum_{r=1}^{q} x_k^{r-1} h_{q-r}(x_1, x_2, \ldots, x_{k-1})} \\
&= h_q(x_1, x_2, \ldots, x_{k-1}) + x_k \underbrace{\sum_{r=1}^{q} x_k^{r-1} h_{q-r}(x_1, x_2, \ldots, x_{k-1})}_{\substack{=h_{q-1}(x_1, x_2, \ldots, x_k) \\ \text{(by (1111))}}} \\
&= h_q(x_1, x_2, \ldots, x_{k-1}) + x_k h_{q-1}(x_1, x_2, \ldots, x_k).
\end{aligned}$$

This proves Lemma 6.54. □

Proof of Lemma 6.55. We first notice that every positive integer q satisfies

$$h_q(x_1, x_2, \ldots, x_0) = 0. \tag{1112}$$

(Here, as usual, $(x_1, x_2, \ldots, x_0)$ stands for the 0-tuple $()$.)

[*Proof of (1112):* Let q be a positive integer. The definition of $h_q(x_1, x_2, \ldots, x_0)$ yields

$$h_q(x_1, x_2, \ldots, x_0) = \sum_{\substack{(a_1, a_2, \ldots, a_0) \in \mathbb{N}^0; \\ a_1 + a_2 + \cdots + a_0 = q}} x_1^{a_1} x_2^{a_2} \cdots x_0^{a_0}. \tag{1113}$$

But there is only one 0-tuple $(a_1, a_2, \ldots, a_0) \in \mathbb{N}^0$, and this 0-tuple satisfies $a_1 + a_2 + \cdots + a_0 = (\text{empty sum}) = 0 \neq q$ (since q is positive). Hence, there exists no $(a_1, a_2, \ldots, a_0) \in \mathbb{N}^0$ satisfying $a_1 + a_2 + \cdots + a_0 = q$. Thus, the sum on the right hand side of (1113) is empty and equals 0. Therefore, (1113) simplifies to $h_q(x_1, x_2, \ldots, x_0) = 0$. This proves (1112).]

We shall now show that

$$\sum_{k=1}^{j} h_{j-k}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) = u^{j-1} \tag{1114}$$

for each $j \in \{1, 2, \ldots, i\}$.

[*Proof of (1114):* We shall prove (1114) by induction over j:
Induction base: We have

$$\sum_{k=1}^{1} h_{1-k}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1}\left(u - x_p\right) = \underbrace{h_{1-1}\left(x_1, x_2, \ldots, x_1\right)}_{\substack{=h_0(x_1,x_2,\ldots,x_1)=1 \\ \text{(by Lemma 6.52 (b)} \\ \text{(applied to 1 instead of } n))}} \underbrace{\prod_{p=1}^{1-1}\left(u - x_p\right)}_{=(\text{empty product})=1}$$
$$= 1 \cdot 1 = 1 = u^{1-1}$$

(since $u^{1-1} = u^0 = 1$). In other words, (1114) holds for $j = 1$. This completes the induction base.

Induction step: Let $m \in \{1, 2, \ldots, i\}$ be such that $m > 1$. Assume that (1114) holds for $j = m - 1$. We must prove that (1114) holds for $j = m$.

We have assumed that (1114) holds for $j = m - 1$. In other words, we have

$$\sum_{k=1}^{m-1} h_{(m-1)-k}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1}\left(u - x_p\right) = u^{(m-1)-1}. \tag{1115}$$

Now, $m > 1 \geq 0$. Moreover,

$$\sum_{k=1}^{m} \underbrace{h_{m-k}\left(x_1, x_2, \ldots, x_k\right)}_{\substack{=h_{m-k}(x_1,x_2,\ldots,x_{k-1})+x_k h_{m-k-1}(x_1,x_2,\ldots,x_k) \\ \text{(by Lemma 6.54 (applied to } q=m-k))}} \prod_{p=1}^{k-1}\left(u - x_p\right)$$
$$= \sum_{k=1}^{m}\left(h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) + x_k h_{m-k-1}\left(x_1, x_2, \ldots, x_k\right)\right) \prod_{p=1}^{k-1}\left(u - x_p\right)$$
$$= \sum_{k=1}^{m} h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) \prod_{p=1}^{k-1}\left(u - x_p\right)$$
$$+ \sum_{k=1}^{m} x_k h_{m-k-1}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1}\left(u - x_p\right). \tag{1116}$$

We have $h_{m-1}\left(x_1, x_2, \ldots, x_{1-1}\right) = 0$ [507]. But $m > 1$. Hence, we can split off the addend for $k = 1$ from the sum $\sum_{k=1}^{m} h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) \prod_{p=1}^{k-1}\left(u - x_p\right)$. We thus

[507] *Proof.* We have $m > 1$, and thus $m - 1 > 0$. Hence, $m - 1$ is a positive integer. Thus, (1112) (applied to $q = m - 1$) yields $h_{m-1}\left(x_1, x_2, \ldots, x_0\right) = 0$. Now, $1 - 1 = 0$ and thus $h_{m-1}\left(x_1, x_2, \ldots, x_{1-1}\right) = h_{m-1}\left(x_1, x_2, \ldots, x_0\right) = 0$. Qed.

obtain

$$\sum_{k=1}^{m} h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) \prod_{p=1}^{k-1} \left(u - x_p\right)$$
$$= \underbrace{h_{m-1}\left(x_1, x_2, \ldots, x_{1-1}\right)}_{=0} \prod_{p=1}^{1-1} \left(u - x_p\right) + \sum_{k=2}^{m} h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) \prod_{p=1}^{k-1} \left(u - x_p\right)$$
$$= \underbrace{0 \prod_{p=1}^{1-1} \left(u - x_p\right)}_{=0} + \sum_{k=2}^{m} h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) \prod_{p=1}^{k-1} \left(u - x_p\right)$$
$$= \sum_{k=2}^{m} h_{m-k}\left(x_1, x_2, \ldots, x_{k-1}\right) \prod_{p=1}^{k-1} \left(u - x_p\right)$$
$$= \sum_{k=1}^{m-1} \underbrace{h_{m-(k+1)}\left(x_1, x_2, \ldots, x_{(k+1)-1}\right)}_{\substack{=h_{m-k-1}(x_1,x_2,\ldots,x_k) \\ \text{(since } m-(k+1)=m-k-1 \\ \text{and } (k+1)-1=k\text{)}}} \underbrace{\prod_{p=1}^{(k+1)-1}}_{\substack{=\prod_{p=1}^{k} \\ \text{(since } (k+1)-1=k\text{)}}} \left(u - x_p\right)$$
(here, we have substituted $k+1$ for k in the sum)
$$= \sum_{k=1}^{m-1} h_{m-k-1}\left(x_1, x_2, \ldots, x_k\right) \underbrace{\prod_{p=1}^{k} \left(u - x_p\right)}_{\substack{=(u-x_k)\prod_{p=1}^{k-1}(u-x_p) \\ \text{(here, we have split off the factor for } p=k \text{ from} \\ \text{the product (since } k\in\{1,2,\ldots,k\}\text{))}}}$$
$$= \sum_{k=1}^{m-1} \underbrace{h_{m-k-1}\left(x_1, x_2, \ldots, x_k\right) \cdot \left(u - x_k\right)}_{=(u-x_k)h_{m-k-1}(x_1,x_2,\ldots,x_k)} \prod_{p=1}^{k-1} \left(u - x_p\right)$$
$$= \sum_{k=1}^{m-1} \left(u - x_k\right) h_{m-k-1}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1} \left(u - x_p\right). \tag{1117}$$

On the other hand, $h_{m-m-1}\left(x_1, x_2, \ldots, x_m\right) = 0$ [508]. But $m > 1$. Hence, we can split off the addend for $k = m$ from the sum $\sum_{k=1}^{m} x_k h_{m-k-1}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1} \left(u - x_p\right)$. We

[508] *Proof.* Clearly, $m - m - 1 = -1$ is a negative integer. Hence, Lemma 6.52 **(a)** (applied to $n = m$ and $k = m - m - 1$) yields $h_{m-m-1}\left(x_1, x_2, \ldots, x_m\right) = 0$. Qed.

thus obtain

$$\begin{aligned}
&\sum_{k=1}^{m} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) \\
&= x_m \underbrace{h_{m-m-1}(x_1, x_2, \ldots, x_m)}_{=0} \prod_{p=1}^{m-1} (u - x_p) + \sum_{k=1}^{m-1} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) \\
&= \underbrace{x_m 0 \prod_{p=1}^{m-1} (u - x_p)}_{=0} + \sum_{k=1}^{m-1} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) \\
&= \sum_{k=1}^{m-1} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) . \qquad (1118)
\end{aligned}$$

Now, (1116) becomes

$$\sum_{k=1}^{m} h_{m-k}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p)$$
$$= \underbrace{\sum_{k=1}^{m} h_{m-k}(x_1, x_2, \ldots, x_{k-1}) \prod_{p=1}^{k-1} (u - x_p)}_{\substack{= \sum_{k=1}^{m-1} (u - x_k) h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) \\ \text{(by (1117))}}}$$
$$+ \underbrace{\sum_{k=1}^{m} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p)}_{\substack{= \sum_{k=1}^{m-1} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p) \\ \text{(by (1118))}}}$$
$$= \sum_{k=1}^{m-1} (u - x_k) h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p)$$
$$+ \sum_{k=1}^{m-1} x_k h_{m-k-1}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p)$$
$$= \sum_{k=1}^{m-1} \underbrace{((u - x_k) + x_k)}_{=u} \underbrace{h_{m-k-1}(x_1, x_2, \ldots, x_k)}_{\substack{= h_{(m-1)-k}(x_1, x_2, \ldots, x_k) \\ \text{(since } m-k-1 = (m-1)-k\text{)}}} \prod_{p=1}^{k-1} (u - x_p)$$
$$= \sum_{k=1}^{m-1} u h_{(m-1)-k}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p)$$
$$= u \underbrace{\sum_{k=1}^{m-1} h_{(m-1)-k}(x_1, x_2, \ldots, x_k) \prod_{p=1}^{k-1} (u - x_p)}_{\substack{= u^{(m-1)-1} \\ \text{(by (1115))}}} = u u^{(m-1)-1} = u^{m-1}.$$

In other words, (1114) holds for $j = m$. This completes the induction step.

Thus, we have proven (1114) by induction.]

Lemma 6.55 now follows from (1114) (applied to $j = i$). □

Proof of Lemma 6.56. For every $(i, j) \in \{1, 2, \ldots, n\}^2$, set

$$a_{i,j} = \prod_{p=1}^{j-1} (x_i - x_p). \tag{1119}$$

Recall that $U = \left(\prod_{p=1}^{i-1} \left(x_j - x_p \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, the definition of U^T yields

$$U^T = \left(\underbrace{\prod_{p=1}^{j-1} \left(x_i - x_p \right)}_{\substack{=a_{i,j} \\ \text{(by (1119))}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

We have $a_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$ [509]. Hence, Exercise 6.3 (applied to U^T instead of A) yields

$$\det \left(U^T \right) = a_{1,1} a_{2,2} \cdots a_{n,n} = \prod_{i=1}^{n} \underbrace{a_{i,i}}_{\substack{= \prod_{p=1}^{i-1} \left(x_i - x_p \right) \\ \text{(by the definition of } a_{i,i})}} = \underbrace{\prod_{i=1}^{n} \prod_{p=1}^{i-1}}_{= \prod_{1 \leq p < i \leq n}} \left(x_i - x_p \right)$$
$$= \prod_{1 \leq p < i \leq n} \left(x_i - x_p \right) = \prod_{1 \leq j < i \leq n} \left(x_i - x_j \right)$$

(here, we have renamed the index (p,i) as (j,i) in the product).

But Exercise 6.4 (applied to $A = U$) yields $\det \left(U^T \right) = \det U$. Hence,

$$\det U = \det \left(U^T \right) = \prod_{1 \leq j < i \leq n} \left(x_i - x_j \right).$$

This proves Lemma 6.56. □

Proof of Lemma 6.58. For every $(i,j) \in \{1,2,\ldots,n\}^2$, set

$$a_{i,j} = h_{i-j} \left(x_1, x_2, \ldots, x_j \right). \tag{1120}$$

[509] *Proof.* Let $(i,j) \in \{1,2,\ldots,n\}^2$ be such that $i < j$. We want to show that $a_{i,j} = 0$.

We have $i < j$, and thus $i \leq j-1$ (since i and j are integers). Thus, $i \in \{1,2,\ldots,j-1\}$. Thus, the product $\prod_{p=1}^{j-1} \left(x_i - x_p \right)$ has a factor for $p = i$. This factor is $x_i - x_i = 0$. Hence, at least one factor of the product $\prod_{p=1}^{j-1} \left(x_i - x_p \right)$ equals 0 (namely, the factor for $p = i$). Thus, the whole product $\prod_{p=1}^{j-1} \left(x_i - x_p \right)$ equals 0 (because if at least one factor of a product equals 0, then the whole product must equal 0). In other words, $\prod_{p=1}^{j-1} \left(x_i - x_p \right) = 0$.

But the definition of $a_{i,j}$ yields $a_{i,j} = \prod_{p=1}^{j-1} \left(x_i - x_p \right) = 0$. Qed.

Then,

$$L = \left(\underbrace{h_{i-j}\left(x_1, x_2, \ldots, x_j\right)}_{\substack{=a_{i,j} \\ \text{(by (1120))}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

We have $a_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$ [510]. Hence, Exercise 6.3 (applied to L instead of A) yields

$$\det L = a_{1,1}a_{2,2}\cdots a_{n,n} = \prod_{i=1}^{n} \underbrace{a_{i,i}}_{\substack{=h_{i-i}(x_1,x_2,\ldots,x_i) \\ \text{(by the definition of } a_{i,i})}} = \prod_{i=1}^{n} \underbrace{h_{i-i}\left(x_1, x_2, \ldots, x_i\right)}_{\substack{=h_0(x_1,x_2,\ldots,x_i) \\ \text{(since } i-i=0)}}$$

$$= \prod_{i=1}^{n} \underbrace{h_0\left(x_1, x_2, \ldots, x_i\right)}_{\substack{=1 \\ \text{(by Lemma 6.52 (b)} \\ \text{(applied to } i \text{ instead of } n))}} = \prod_{i=1}^{n} 1 = 1.$$

This proves Lemma 6.58. ☐

Proof of Lemma 6.60. If $i \in \mathbb{Z}$ and $k \in \{1,2,\ldots,n\}$ are such that $k > i$, then

$$h_{i-k}\left(x_1, x_2, \ldots, x_k\right) = 0 \tag{1121}$$

[511].

Now, for every $i \in \{1,2,\ldots,n\}$ and $u \in \mathbb{K}$, we have

$$\sum_{k=1}^{n} h_{i-k}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1} \left(u - x_p\right)$$

$$= \underbrace{\sum_{k=1}^{i} h_{i-k}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1} \left(u - x_p\right)}_{\substack{=u^{i-1} \\ \text{(by Lemma 6.55)}}} + \sum_{k=i+1}^{n} \underbrace{h_{i-k}\left(x_1, x_2, \ldots, x_k\right)}_{\substack{=0 \\ \text{(by (1121))} \\ \text{(since } k \geq i+1 > i))}} \prod_{p=1}^{k-1} \left(u - x_p\right)$$

$$\left(\text{since } 1 \leq i \leq n \text{ (since } i \in \{1,2,\ldots,n\}\text{)}\right)$$

$$= u^{i-1} + \underbrace{\sum_{k=i+1}^{n} 0 \prod_{p=1}^{k-1} \left(u - x_p\right)}_{=0} = u^{i-1}. \tag{1122}$$

[510] *Proof.* Let $(i,j) \in \{1,2,\ldots,n\}^2$ be such that $i < j$. We want to show that $a_{i,j} = 0$.
We have $i - j < 0$ (since $i < j$). Thus, $i - j$ is a negative integer. Hence, Lemma 6.52 **(a)** (applied to j and $i - j$ instead of n and k) yields $h_{i-j}\left(x_1, x_2, \ldots, x_j\right) = 0$. But the definition of $a_{i,j}$ yields $a_{i,j} = h_{i-j}\left(x_1, x_2, \ldots, x_j\right) = 0$. Qed.

[511] *Proof of (1121):* Let $i \in \mathbb{Z}$ and $k \in \{1,2,\ldots,n\}$ be such that $k > i$. Then, $i - k < 0$ (since $k > i$). Thus, $i - k$ is a negative integer. Hence, Lemma 6.52 **(a)** (applied to k and $i - k$ instead of n and k) yields $h_{i-k}\left(x_1, x_2, \ldots, x_k\right) = 0$. This proves (1121).

But recall that $L = \left(h_{i-j}\left(x_1, x_2, \ldots, x_j\right)\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ and $U = \left(\prod_{p=1}^{i-1}\left(x_j - x_p\right)\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, the definition of the product LU yields

$$LU = \left(\underbrace{\sum_{k=1}^{n} h_{i-k}\left(x_1, x_2, \ldots, x_k\right) \prod_{p=1}^{k-1}\left(x_j - x_p\right)}_{\substack{=x_j^{i-1} \\ \text{(by (1122) (applied to } u=x_j\text{))}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(x_j^{i-1}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

This proves Lemma 6.60. □

Solution to Exercise 6.15. We have proven Lemma 6.52, Lemma 6.53, Lemma 6.54, Lemma 6.55, Lemma 6.56, Lemma 6.58 and Lemma 6.60. Thus, Exercise 6.15 is solved. □

7.83. Solution to Exercise 6.16

First solution to Exercise 6.16. Our solution will imitate our First proof of Theorem 6.46 **(a)** (but it will involve some additional complications).

For every $u \in \{0, 1, \ldots, n\}$ and $(i, j) \in \{1, 2, \ldots, u\}^2$, define $a_{i,j,u} \in \mathbb{K}$ by

$$a_{i,j,u} = \begin{cases} x_i^{u-j}, & \text{if } j > 1; \\ x_i^u, & \text{if } j = 1 \end{cases}.$$

For every $u \in \{1, 2, \ldots, n\}$, let A_u be the $u \times u$-matrix $\left(a_{i,j,u}\right)_{1 \leq i \leq u,\ 1 \leq j \leq u}$. Then, our goal is to prove that $\det\left(A_n\right) = \left(x_1 + x_2 + \cdots + x_n\right) \prod_{1 \leq i < j \leq n}\left(x_i - x_j\right)$ (because A_n is precisely the matrix $\left(\begin{cases} x_i^{n-j}, & \text{if } j > 1; \\ x_i^n, & \text{if } j = 1 \end{cases}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ which appears in the statement of the exercise).

Now, let us show that

$$\det\left(A_u\right) = \left(x_1 + x_2 + \cdots + x_u\right) \prod_{1 \leq i < j \leq u}\left(x_i - x_j\right) \tag{1123}$$

for every $u \in \{1, 2, \ldots, n\}$.

[*Proof of (1123):* We will prove (1123) by induction over u:

Induction base: The definition of A_1 yields $A_1 = \left(x_1^1 \right) = \left(x_1 \right)$ and thus $\det(A_1) = x_1$. Compared with

$$\underbrace{(x_1 + x_2 + \cdots + x_1)}_{= x_1} \underbrace{\prod_{1 \leq i < j \leq 1} \left(x_i - x_j\right)}_{= (\text{empty product}) = 1} = x_1,$$

this yields $\det(A_1) = (x_1 + x_2 + \cdots + x_1) \prod_{1 \leq i < j \leq 1} \left(x_i - x_j\right)$. In other words, (1123) holds for $u = 1$. The induction base is thus complete.

Induction step: Let $U \in \{2, 3, \ldots, n\}$. Assume that (1123) holds for $u = U - 1$. We need to prove that (1123) holds for $u = U$.

Recall that $A_U = \left(a_{i,j,U}\right)_{1 \leq i \leq U,\ 1 \leq j \leq U}$ (by the definition of A_U). The matrix A_U looks as follows:

$$A_U = \begin{pmatrix} x_1^U & x_1^{U-2} & x_1^{U-3} & \cdots & x_1 & 1 \\ x_2^U & x_2^{U-2} & x_2^{U-3} & \cdots & x_2 & 1 \\ x_3^U & x_3^{U-2} & x_3^{U-3} & \cdots & x_3 & 1 \\ x_4^U & x_4^{U-2} & x_4^{U-3} & \cdots & x_4 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_U^U & x_U^{U-2} & x_U^{U-3} & \cdots & x_U & 1 \end{pmatrix}.$$

For every $(i, j) \in \{1, 2, \ldots, U\}^2$, define $b_{i,j} \in \mathbb{K}$ by

$$b_{i,j} = \begin{cases} x_i^U - x_U^2 x_i^{U-2}, & \text{if } j = 1; \\ x_i^{U-j} - x_U x_i^{U-j-1}, & \text{if } 1 < j < U; \\ 1, & \text{if } j = U \end{cases}.$$

Let B be the $U \times U$-matrix $\left(b_{i,j}\right)_{1 \leq i \leq U,\ 1 \leq j \leq U}$. Here is how B looks like:

$$B = \begin{pmatrix} x_1^U - x_U^2 x_1^{U-2} & x_1^{U-2} - x_U x_1^{U-3} & x_1^{U-3} - x_U x_1^{U-4} & \cdots & x_1 - x_U & 1 \\ x_2^U - x_U^2 x_2^{U-2} & x_2^{U-2} - x_U x_2^{U-3} & x_2^{U-3} - x_U x_2^{U-4} & \cdots & x_2 - x_U & 1 \\ x_3^U - x_U^2 x_3^{U-2} & x_3^{U-2} - x_U x_3^{U-3} & x_3^{U-3} - x_U x_3^{U-4} & \cdots & x_3 - x_U & 1 \\ x_4^U - x_U^2 x_4^{U-2} & x_4^{U-2} - x_U x_4^{U-3} & x_4^{U-3} - x_U x_4^{U-4} & \cdots & x_4 - x_U & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_U^U - x_U^2 x_U^{U-2} & x_U^{U-2} - x_U x_U^{U-3} & x_U^{U-3} - x_U x_U^{U-4} & \cdots & x_U - x_U & 1 \end{pmatrix}.$$

We claim that $\det B = \det(A_U)$. Indeed, here are two ways to prove this:

First proof of $\det B = \det(A_U)$*:* Exercise 6.8 **(b)** shows that the determinant of a $U \times U$-matrix does not change if we subtract a multiple of one of its columns from another column. Now, let us do the following steps (in this order):

- subtract x_U^2 times the 2-nd column of A_U from the 1-st column;
- subtract x_U times the 3-rd column of the resulting matrix from the 2-nd column;
- subtract x_U times the 4-th column of the resulting matrix from the 3-rd column;
- and so on, all the way until we finally subtract x_U times the U-th column of the matrix from the $(U-1)$-st column.

Yes, you are reading this right: At the first step we subtract x_U^2 times (not x_U times) the 2-nd column from the 1-st column; but at all further steps, we subtract x_U times a column from another. Having done all this, the resulting matrix is B (according to our definition of B). Thus, $\det B = \det(A_U)$ (since our subtractions never change the determinant). This proves $\det B = \det(A_U)$.

Second proof of $\det B = \det(A_U)$*:* Here is another way to prove that $\det B = \det(A_U)$, with some less handwaving.

For every $(i,j) \in \{1,2,\ldots,U\}^2$, we define $c_{i,j} \in \mathbb{K}$ by

$$c_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ -x_U^2, & \text{if } i = j+1 \text{ and } j = 1; \\ -x_U, & \text{if } i = j+1 \text{ and } j > 1; \\ 0, & \text{otherwise} \end{cases}.$$

Let C be the $U \times U$-matrix $(c_{i,j})_{1 \leq i \leq U,\ 1 \leq j \leq U}$. Here is how C looks like:

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -x_U^2 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -x_U & 1 & \cdots & 0 & 0 \\ 0 & 0 & -x_U & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -x_U & 1 \end{pmatrix},$$

where the only $-x_U^2$ is in the $(2,1)$-th cell.

The matrix C is lower-triangular, and thus Exercise 6.3 shows that its determinant is $\det C = \underbrace{c_{1,1}}_{=1} \underbrace{c_{2,2}}_{=1} \cdots \underbrace{c_{U,U}}_{=1} = 1$.

On the other hand, it is easy to see that $B = A_U C$ (check this!). Thus, Theorem 6.23 yields $\det B = \det(A_U) \cdot \underbrace{\det C}_{=1} = \det(A_U)$. So we have proven $\det B = \det(A_U)$ again.

Next, we observe that for every $j \in \{1, 2, \ldots, U-1\}$, we have

$$\begin{aligned} b_{U,j} &= \begin{cases} x_U^U - x_U^2 x_U^{U-2}, & \text{if } j = 1; \\ x_U^{U-j} - x_U x_U^{U-j-1}, & \text{if } 1 < j < U; \\ 1, & \text{if } j = U \end{cases} \qquad \left(\text{by the definition of } b_{U,j}\right) \\ &= \begin{cases} x_U^U - x_U^2 x_U^{U-2}, & \text{if } j = 1; \\ x_U^{U-j} - x_U x_U^{U-j-1}, & \text{if } 1 < j < U \end{cases} \\ &\qquad (\text{since } j < U \text{ (since } j \in \{1, 2, \ldots, U-1\})) \\ &= \begin{cases} x_U^U - x_U^U, & \text{if } j = 1; \\ x_U^{U-j} - x_U^{U-j}, & \text{if } 1 < j < U \end{cases} = \begin{cases} 0, & \text{if } j = 1; \\ 0, & \text{if } 1 < j < U \end{cases} = 0. \end{aligned}$$

Hence, Theorem 6.43 (applied to U, B and $b_{i,j}$ instead of n, A and $a_{i,j}$) yields

$$\det B = b_{U,U} \cdot \det\left(\left(b_{i,j}\right)_{1 \le i \le U-1,\ 1 \le j \le U-1}\right). \tag{1124}$$

Let B' denote the $(U-1) \times (U-1)$-matrix $\left(b_{i,j}\right)_{1 \le i \le U-1,\ 1 \le j \le U-1}$.

The definition of $b_{U,U}$ yields

$$\begin{aligned} b_{U,U} &= \begin{cases} x_U^U - x_U^2 x_U^{U-2}, & \text{if } U = 1; \\ x_U^{U-U} - x_U x_U^{U-U-1}, & \text{if } 1 < U < U; \\ 1, & \text{if } U = U \end{cases} \qquad \left(\text{by the definition of } b_{U,U}\right) \\ &= 1 \qquad (\text{since } U = U). \end{aligned}$$

Thus, (1124) becomes

$$\det B = \underbrace{b_{U,U}}_{=1} \cdot \det\left(\underbrace{\left(b_{i,j}\right)_{1 \le i \le U-1,\ 1 \le j \le U-1}}_{=B'}\right) = \det\left(B'\right).$$

Compared with $\det B = \det\left(A_U\right)$, this yields

$$\det\left(A_U\right) = \det\left(B'\right). \tag{1125}$$

Now, let us take a closer look at B'. Indeed, every $(i, j) \in \{1, 2, \ldots, U-1\}^2$

satisfies

$$\begin{aligned}
b_{i,j} &= \begin{cases} x_i^U - x_U^2 x_i^{U-2}, & \text{if } j=1; \\ x_i^{U-j} - x_U x_i^{U-j-1}, & \text{if } 1<j<U; \\ 1, & \text{if } j=U \end{cases} \qquad \left(\text{by the definition of } b_{i,j}\right) \\
&= \begin{cases} x_i^U - x_U^2 x_i^{U-2}, & \text{if } j=1; \\ x_i^{U-j} - x_U x_i^{U-j-1}, & \text{if } 1<j<U \end{cases} \\
&\qquad \left(\begin{array}{c} \text{since } j<U \text{ (since } j\in\{1,2,\ldots,U-1\} \\ \text{(since } (i,j)\in\{1,2,\ldots,U-1\}^2)) \end{array}\right) \\
&= \begin{cases} \left(x_i^2 - x_U^2\right) x_i^{U-2}, & \text{if } j=1; \\ (x_i - x_U) x_i^{(U-1)-j}, & \text{if } 1<j<U \end{cases} = \begin{cases} (x_i - x_U)(x_i + x_U) x_i^{U-2}, & \text{if } j=1; \\ (x_i - x_U) x_i^{(U-1)-j}, & \text{if } 1<j<U \end{cases} \\
&= (x_i - x_U) \begin{cases} (x_i + x_U) x_i^{U-2}, & \text{if } j=1; \\ x_i^{(U-1)-j}, & \text{if } 1<j<U \end{cases} \\
&= (x_i - x_U) \begin{cases} (x_i + x_U) x_i^{U-2}, & \text{if } j=1; \\ x_i^{(U-1)-j}, & \text{if } j>1 \end{cases} \qquad (1126)
\end{aligned}$$

(since $1<j<U$ is equivalent to $j>1$).

For every $(i,j)\in\{1,2,\ldots,U-1\}^2$, we define $g_{i,j}\in\mathbb{K}$ by

$$g_{i,j} = \begin{cases} (x_i + x_U) x_i^{U-2}, & \text{if } j=1; \\ x_i^{(U-1)-j}, & \text{if } j>1 \end{cases}.$$

Let G be the $(U-1)\times(U-1)$-matrix $\left(g_{i,j}\right)_{1\le i\le U-1,\ 1\le j\le U-1}$. Here is how G looks like:

$$G = \begin{pmatrix} (x_1 + x_U) x_1^{U-2} & x_1^{U-3} & x_1^{U-4} & \cdots & x_1 & 1 \\ (x_2 + x_U) x_2^{U-2} & x_2^{U-3} & x_2^{U-4} & \cdots & x_2 & 1 \\ (x_3 + x_U) x_3^{U-2} & x_3^{U-3} & x_3^{U-4} & \cdots & x_3 & 1 \\ (x_4 + x_U) x_4^{U-2} & x_4^{U-3} & x_4^{U-4} & \cdots & x_4 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ (x_{U-1} + x_U) x_{U-1}^{U-2} & x_{U-1}^{U-3} & x_{U-1}^{U-4} & \cdots & x_{U-1} & 1 \end{pmatrix}.$$

Now, (1126) becomes

$$b_{i,j} = (x_i - x_U) \underbrace{\begin{cases} (x_i + x_U) x_i^{U-2}, & \text{if } j=1; \\ x_i^{(U-1)-j}, & \text{if } j>1 \end{cases}}_{=g_{i,j}} = (x_i - x_U) g_{i,j} \qquad (1127)$$

for every $(i,j) \in \{1,2,\ldots,U-1\}^2$. Hence,

$$B' = \left(\underbrace{b_{i,j}}_{\substack{=(x_i - x_U) g_{i,j} \\ \text{(by (1127))}}} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1} = \left((x_i - x_U)\, g_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}. \qquad (1128)$$

On the other hand, the definition of G yields

$$G = \left(g_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}. \qquad (1129)$$

Now, we claim that

$$\det \left(B' \right) = \det G \cdot \prod_{i=1}^{U-1} (x_i - x_U). \qquad (1130)$$

This can be proven similarly to how we proved (409) back in our First proof of Theorem 6.46. (Of course, this time, G plays the role of A_{U-1}.) Now, (1125) becomes

$$\det (A_U) = \det \left(B' \right) = \det G \cdot \prod_{i=1}^{U-1} (x_i - x_U). \qquad (1131)$$

Now, we want to compute $\det G$. (This part is harder than the analogous part of our First proof of Theorem 6.46, because back then the role of G was played by A_{U-1}, and we knew $\det (A_{U-1})$ directly from our induction hypothesis.)

For every $(i,j) \in \{1,2,\ldots,U-1\}^2$, we define two elements $p_{i,j} \in \mathbb{K}$ and $q_{i,j} \in \mathbb{K}$ by

$$p_{i,j} = \begin{cases} x_i^{U-1}, & \text{if } j = 1; \\ x_i^{(U-1)-j}, & \text{if } j > 1 \end{cases}$$

and

$$q_{i,j} = \begin{cases} x_U x_i^{U-2}, & \text{if } j = 1; \\ x_i^{(U-1)-j}, & \text{if } j > 1 \end{cases}.$$

We let P be the $(U-1) \times (U-1)$-matrix $\left(p_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}$, and we let Q be the $(U-1) \times (U-1)$-matrix $\left(q_{i,j} \right)_{1 \leq i \leq U-1,\ 1 \leq j \leq U-1}$. We now make the following observations:

- The columns of the matrix Q equal the corresponding columns of P except (perhaps) the 1-st column. The matrix G is the $(U-1) \times (U-1)$-matrix obtained from P by adding the 1-st column of Q to the 1-st column of P. Thus, Exercise 6.7 **(j)** (applied to $U-1$, 1, P, Q and G instead of n, k, A, A' and B) yields $\det G = \det P + \det Q$.

- We have $P = \left(p_{i,j}\right)_{1\leq i\leq U-1,\ 1\leq j\leq U-1}$ and $A_{U-1} = \left(a_{i,j,U-1}\right)_{1\leq i\leq U-1,\ 1\leq j\leq U-1}$. But a quick look at the definitions reveals that

$$p_{i,j} = \begin{cases} x_i^{U-1}, & \text{if } j=1; \\ x_i^{(U-1)-j}, & \text{if } j>1 \end{cases} = \begin{cases} x_i^{(U-1)-j}, & \text{if } j>1; \\ x_i^{U-1}, & \text{if } j=1 \end{cases} = a_{i,j,U-1}$$

for all $(i,j) \in \{1,2,\ldots,U-1\}^2$. Hence, $P = A_{U-1}$.

- The matrix Q is obtained from the matrix $\left(x_i^{(U-1)-j}\right)_{1\leq i\leq U-1,\ 1\leq j\leq U-1}$ by multiplying its 1-st column by x_U. Hence, Exercise 6.7 **(h)** (applied to $U-1$, x_U, 1, $\left(x_i^{(U-1)-j}\right)_{1\leq i\leq U-1,\ 1\leq j\leq U-1}$ and Q instead of n, λ, k, A and B) yields

$$\det Q = x_U \underbrace{\det\left(\left(x_i^{(U-1)-j}\right)_{1\leq i\leq U-1,\ 1\leq j\leq U-1}\right)}_{\substack{=\prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right) \\ \text{(by Theorem 6.46 \textbf{(a)}, applied to } n=U-1\text{)}}} = x_U \prod_{1\leq i<j\leq U-1}\left(x_i - x_j\right).$$

Combining this, we obtain

$$\begin{aligned}
\det G &= \det \underbrace{P}_{=A_{U-1}} + \underbrace{\det Q}_{=x_U \prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right)} \\
&= \underbrace{\det\left(A_{U-1}\right)}_{\substack{=\left(x_1+x_2+\cdots+x_{U-1}\right)\prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right) \\ \text{(since we assumed that (1123) holds for } u=U-1\text{)}}} + x_U \prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right) \\
&= \left(x_1+x_2+\cdots+x_{U-1}\right)\prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right) + x_U \prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right) \\
&= \underbrace{\left(\left(x_1+x_2+\cdots+x_{U-1}\right)+x_U\right)}_{=x_1+x_2+\cdots+x_U}\prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right) \\
&= \left(x_1+x_2+\cdots+x_U\right)\prod_{1\leq i<j\leq U-1}\left(x_i-x_j\right).
\end{aligned}$$

Hence, (1131) yields

$$\begin{aligned}\det\left(A_U\right) &= \underbrace{\det G}_{=\left(x_1+x_2+\cdots+x_U\right)\prod\limits_{1\leq i<j\leq U-1}\left(x_i-x_j\right)} \cdot \prod_{i=1}^{U-1}\left(x_i-x_U\right) \\ &= \left(x_1+x_2+\cdots+x_U\right) \underbrace{\prod_{1\leq i<j\leq U-1}}_{=\prod\limits_{j=1}^{U-1}\prod\limits_{i=1}^{j-1}} \left(x_i-x_j\right) \cdot \prod_{i=1}^{U-1}\left(x_i-x_U\right) \\ &= \left(x_1+x_2+\cdots+x_U\right)\left(\prod_{j=1}^{U-1}\prod_{i=1}^{j-1}\left(x_i-x_j\right)\right) \cdot \prod_{i=1}^{U-1}\left(x_i-x_U\right).\end{aligned}$$

Compared with

$$\begin{aligned}&\left(x_1+x_2+\cdots+x_U\right) \underbrace{\prod_{1\leq i<j\leq U}}_{=\prod\limits_{j=1}^{U}\prod\limits_{i=1}^{j-1}} \left(x_i-x_j\right) \\ &= \left(x_1+x_2+\cdots+x_U\right)\prod_{j=1}^{U}\prod_{i=1}^{j-1}\left(x_i-x_j\right) \\ &= \left(x_1+x_2+\cdots+x_U\right)\left(\prod_{j=1}^{U-1}\prod_{i=1}^{j-1}\left(x_i-x_j\right)\right) \cdot \prod_{i=1}^{U-1}\left(x_i-x_U\right) \\ &\qquad\qquad \left(\text{here, we have split off the factor for } j=U \text{ from the product}\right),\end{aligned}$$

this yields $\det\left(A_U\right) = \left(x_1+x_2+\cdots+x_U\right) \prod\limits_{1\leq i<j\leq U}\left(x_i-x_j\right)$. In other words, (1123) holds for $u=U$. This completes the induction step.

Now, (1123) is proven by induction.]

Hence, we can apply (1123) to $u=n$. As the result, we obtain $\det\left(A_n\right) = \left(x_1+x_2+\cdots+x_n\right) \prod\limits_{1\leq i<j\leq n}\left(x_i-x_j\right)$. Since $A_n = \left(\begin{cases} x_i^{n-j}, & \text{if } j>1; \\ x_i^n, & \text{if } j=1 \end{cases}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of A_n), this rewrites as

$$\det\left(\left(\begin{cases} x_i^{n-j}, & \text{if } j>1; \\ x_i^n, & \text{if } j=1 \end{cases}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \left(x_1+x_2+\cdots+x_n\right) \prod_{1\leq i<j\leq n}\left(x_i-x_j\right).$$

This solves Exercise 6.16. □

We shall give a second solution for Exercise 6.16 later (in Section 7.88).

7.84. Solution to Exercise 6.17

We shall first sketch short solutions to parts **(a)** and **(b)** of Exercise 6.17. Then, we will show a solution to part **(c)** (which is a more elaborate and subtler version of our solution to part **(b)**), and finally derive parts **(a)** and **(b)** from part **(c)**. Thus, parts **(a)** and **(b)** of Exercise 6.17 will be solved twice (though the solutions cannot really be called different).

*Solution sketch to parts **(a)** and **(b)** of Exercise 6.17.* **(a)** Let $m \in \{0, 1, \ldots, n-2\}$. Thus, $m + 1 < n$.

Define an $n \times (m+1)$-matrix B by

$$B = \left(\binom{m}{j-1} x_i^{j-1} \right)_{1 \leq i \leq n,\ 1 \leq j \leq m+1}.$$

Define an $(m+1) \times n$-matrix C by

$$C = \left(y_j^{m-(i-1)} \right)_{1 \leq i \leq m+1,\ 1 \leq j \leq n}.$$

Then, the definition of the product of two matrices shows that

$$BC = \left(\sum_{k=1}^{m+1} \binom{m}{k-1} x_i^{k-1} y_j^{m-(k-1)} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Since every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfies

$$\begin{aligned} &\sum_{k=1}^{m+1} \binom{m}{k-1} x_i^{k-1} y_j^{m-(k-1)} \\ &= \sum_{k=0}^{m} \binom{m}{k} x_i^k y_j^{m-k} \qquad \text{(here, we have substituted } k \text{ for } k-1 \text{ in the sum)} \\ &= (x_i + y_j)^m \qquad \left(\begin{array}{c} \text{since the binomial formula yields} \\ (x_i + y_j)^m = \sum_{k=0}^{m} \binom{m}{k} x_i^k y_j^{m-k} \end{array} \right), \end{aligned}$$

this rewrites as

$$BC = \left((x_i + y_j)^m \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

But recall that $m + 1 < n$. Hence, (379) (applied to $m+1$, B and C instead of m, A and B) shows that $\det(BC) = 0$. Since $BC = \left((x_i + y_j)^m \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, this rewrites as $\det\left(\left((x_i + y_j)^m \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) = 0$. This solves Exercise 6.17 **(a)**.

(b) Define an $n \times n$-matrix B by

$$B = \left(\binom{n-1}{j-1} x_i^{j-1} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Define an $n \times n$-matrix C by

$$C = \left(y_j^{n-i}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Then, the definition of the product of two matrices shows that

$$BC = \left(\sum_{k=1}^{n} \binom{n-1}{k-1} x_i^{k-1} y_j^{n-k}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Since every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfies

$$\sum_{k=1}^{n} \binom{n-1}{k-1} x_i^{k-1} \underbrace{y_j^{n-k}}_{= y_j^{(n-1)-(k-1)}} = \sum_{k=1}^{n} \binom{n-1}{k-1} x_i^{k-1} y_j^{(n-1)-(k-1)}$$

$$= \sum_{k=0}^{n-1} \binom{n-1}{k} x_i^k y_j^{(n-1)-k} \qquad \text{(here, we have substituted } k \text{ for } k-1 \text{ in the sum)}$$

$$= \left(x_i + y_j\right)^{n-1} \qquad \left(\begin{array}{c}\text{since the binomial formula yields} \\ \left(x_i + y_j\right)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} x_i^k y_j^{(n-1)-k}\end{array}\right),$$

this rewrites as

$$BC = \left(\left(x_i + y_j\right)^{n-1}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Theorem 6.23 shows that $\det(BC) = \det B \cdot \det C$. But what are $\det B$ and $\det C$?

Finding $\det C$ is easy: We have $C = \left(y_j^{n-i}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ and thus

$$\det C = \det\left(\left(y_j^{n-i}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = \prod_{1 \leq i < j \leq n} \left(y_i - y_j\right)$$

(by Theorem 6.46 **(b)**, applied to y_k instead of x_k).

To find $\det B$, we first observe that $\det\left(\left(x_i^{j-1}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = \prod_{1 \leq j < i \leq n} \left(x_i - x_j\right)$ (by Theorem 6.46 **(c)**). But $B = \left(\binom{n-1}{j-1} x_i^{j-1}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. In other words, the matrix B is obtained from the matrix $\left(x_i^{j-1}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ by multiplying the whole j-th column with $\binom{n-1}{j-1}$ for every $j \in \{1,2,\ldots,n\}$. Therefore, the determinant $\det B$ is obtained by successively multiplying $\det\left(\left(x_i^{j-1}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right)$ with $\binom{n-1}{j-1}$ for every $j \in \{1,2,\ldots,n\}$ (since Exercise 6.7 **(h)** tells us that multiplying a single column of a square matrix by a scalar λ results in the determinant getting multiplied by λ). In other words,

$$\det B = \underbrace{\det\left(\left(x_i^{j-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{\substack{=\prod_{1\leq j<i\leq n}(x_i-x_j) \\ =\prod_{1\leq i<j\leq n}(x_j-x_i) \\ \text{(here, we renamed the} \\ \text{index } (j,i) \text{ as } (i,j))}} \cdot \underbrace{\prod_{j=1}^{n}\binom{n-1}{j-1}}_{\substack{=\prod_{k=0}^{n-1}\binom{n-1}{k} \\ \text{(here, we substituted } k \text{ for } j-1)}}$$
$$= \left(\prod_{1\leq i<j\leq n}(x_j-x_i)\right)\cdot\left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right).$$

Now,

$$\det(BC) = \underbrace{\det B}_{=\left(\prod_{1\leq i<j\leq n}(x_j-x_i)\right)\cdot\left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)} \cdot \underbrace{\det C}_{=\prod_{1\leq i<j\leq n}(y_i-y_j)}$$
$$= \left(\prod_{1\leq i<j\leq n}(x_j-x_i)\right)\cdot\left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\cdot\left(\prod_{1\leq i<j\leq n}(y_i-y_j)\right)$$
$$= \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\cdot\underbrace{\left(\prod_{1\leq i<j\leq n}(x_j-x_i)\right)\cdot\left(\prod_{1\leq i<j\leq n}(y_i-y_j)\right)}_{=\prod_{1\leq i<j\leq n}((x_j-x_i)(y_i-y_j))}$$
$$= \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\cdot\prod_{1\leq i<j\leq n}\underbrace{((x_j-x_i)(y_i-y_j))}_{=(x_i-x_j)(y_j-y_i)}$$
$$= \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\cdot\underbrace{\prod_{1\leq i<j\leq n}((x_i-x_j)(y_j-y_i))}_{=\left(\prod_{1\leq i<j\leq n}(x_i-x_j)\right)\left(\prod_{1\leq i<j\leq n}(y_j-y_i)\right)}$$
$$= \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\cdot\left(\prod_{1\leq i<j\leq n}(x_i-x_j)\right)\left(\prod_{1\leq i<j\leq n}(y_j-y_i)\right).$$

Since $BC = \left(\left(x_i + y_j \right)^{n-1} \right)_{1 \leq i \leq n, \ 1 \leq j \leq n}$, this rewrites as

$$\det \left(\left(\left(x_i + y_j \right)^{n-1} \right)_{1 \leq i \leq n, \ 1 \leq j \leq n} \right) = \left(\prod_{k=0}^{n-1} \dbinom{n-1}{k} \right) \cdot \left(\prod_{1 \leq i < j \leq n} \left(x_i - x_j \right) \right) \left(\prod_{1 \leq i < j \leq n} \left(y_j - y_i \right) \right).$$

This solves Exercise 6.17 **(b)**. □

Now, as promised, we shall start from scratch, and solve Exercise 6.17 **(c)** first, and then solve Exercise 6.17 **(a)** and **(b)** again using Exercise 6.17 **(c)**.

Solution to Exercise 6.17. **(c)** We extend the n-tuple $(p_0, p_1, \ldots, p_{n-1})$ to an infinite sequence $(p_0, p_1, p_2, \ldots) \in \mathbb{K}^\infty$ by setting

$$p_\ell = 0 \qquad \text{for every } \ell \in \{n, n+1, n+2, \ldots\}. \tag{1132}$$

Next, we define three $n \times n$-matrices B, C and D:

- Define an $n \times n$-matrix B by

  $$B = \left(x_i^{n-j} \right)_{1 \leq i \leq n, \ 1 \leq j \leq n}.$$

  Then,

  $$\det \underbrace{B}_{= \left(x_i^{n-j} \right)_{1 \leq i \leq n, \ 1 \leq j \leq n}} = \det \left(\left(x_i^{n-j} \right)_{1 \leq i \leq n, \ 1 \leq j \leq n} \right) = \prod_{1 \leq i < j \leq n} \left(x_i - x_j \right) \tag{1133}$$

  (by Theorem 6.46 **(a)**).

- For every $(i, j) \in \{1, 2, \ldots, n\}^2$, the element $\dbinom{n-1-i+j}{j-1} p_{n-1-i+j}$ of $\mathbb{K}$ is well-defined[512]. Thus, for every $(i, j) \in \{1, 2, \ldots, n\}^2$, we can define an element $c_{i,j} \in \mathbb{K}$ by $c_{i,j} = \dbinom{n-1-i+j}{j-1} p_{n-1-i+j}$. Consider these elements $c_{i,j}$. Define an $n \times n$-matrix C by

  $$C = \left(c_{i,j} \right)_{1 \leq i \leq n, \ 1 \leq j \leq n}.$$

[512] *Proof.* Let $(i, j) \in \{1, 2, \ldots, n\}^2$. Thus, $j - 1 \geq 0$. Hence, the binomial coefficient $\dbinom{n-1-i+j}{j-1} \in \mathbb{Z}$ is well-defined. Moreover, $n - 1 - \underbrace{i}_{\leq n} + \underbrace{j}_{\geq 1} \geq n - 1 - n + 1 = 0$, so that $n - 1 - i + j \in \mathbb{N}$. Thus, $p_{n-1-i+j} \in \mathbb{K}$ is well-defined (since we have an infinite sequence $(p_0, p_1, p_2, \ldots) \in \mathbb{K}^\infty$). Hence, the element $\dbinom{n-1-i+j}{j-1} p_{n-1-i+j}$ of $\mathbb{K}$ is well-defined. Qed.

Consider this matrix C. We have $c_{i,j}=0$ for every $(i,j)\in\{1,2,\ldots,n\}^2$ satisfying $i<j$ [513]. Hence, Exercise 6.3 (applied to C and $\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ instead of A and $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$) shows that

$$\begin{aligned}\det C &= c_{1,1}c_{2,2}\cdots c_{n,n}=\prod_{i=1}^{n}\underbrace{c_{i,i}}_{\substack{=\dbinom{n-1-i+i}{i-1}p_{n-1-i+i}\\ \text{(by the definition of } c_{i,i})}}\\ &=\prod_{i=1}^{n}\left(\underbrace{\dbinom{n-1-i+i}{i-1}}_{=\dbinom{n-1}{i-1}}\underbrace{p_{n-1-i+i}}_{=p_{n-1}}\right)\\ &=\prod_{i=1}^{n}\left(\dbinom{n-1}{i-1}p_{n-1}\right)=\left(\prod_{i=1}^{n}\dbinom{n-1}{i-1}\right)p_{n-1}^{n}\\ &=p_{n-1}^{n}\left(\prod_{i=1}^{n}\dbinom{n-1}{i-1}\right)=p_{n-1}^{n}\left(\prod_{k=0}^{n-1}\dbinom{n-1}{k}\right)\end{aligned}\tag{1134}$$

(here, we have substituted k for $i-1$ in the product).

- Define an $n\times n$-matrix D by

$$D=\left(y_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Then,

$$\begin{aligned}\det\underbrace{D}_{=\left(y_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}} &=\det\left(\left(y_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)=\prod_{1\leq j<i\leq n}\left(y_i-y_j\right)\\ &\qquad\text{(by Theorem 6.46 **(d)**, applied to } y_k \text{ instead of } x_k)\\ &=\prod_{1\leq i<j\leq n}\left(y_j-y_i\right)\end{aligned}\tag{1135}$$

(here, we have renamed the index (j,i) as (i,j) in the product).

[513] *Proof.* Let $(i,j)\in\{1,2,\ldots,n\}^2$ be such that $i<j$. From $i<j$, we obtain $i\leq j-1$ (since i and j are integers). Thus, $n-1-\underbrace{i}_{\leq j-1}+j\geq n-1-(j-1)+j=n$. In other words, $n-1-i+j\in\{n,n+1,n+2,\ldots\}$. Hence, $p_{n-1-i+j}=0$ (by (1132), applied to $\ell=n-1-i+j$). Hence, $c_{i,j}=\dbinom{n-1-i+j}{j-1}\underbrace{p_{n-1-i+j}}_{=0}=0$, qed.

Our goal is to prove that $\left(P\left(x_i+y_j\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}=BCD$. Once this is done, we will be able to compute $\det\left(\left(P\left(x_i+y_j\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$ by applying Theorem 6.23 twice.

We have $C=\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $D=\left(y_j^{i-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence, the definition of the product of two matrices shows that

$$CD=\left(\sum_{k=1}^{n}c_{i,k}y_j^{k-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}. \tag{1136}$$

But every $(i,j)\in\{1,2,\ldots,n\}^2$ satisfies

$$\begin{aligned}\sum_{k=1}^{n}c_{i,k}y_j^{k-1}&=\sum_{\ell=0}^{n-1}\underbrace{c_{i,\ell+1}}_{\substack{=\dbinom{n-1-i+(\ell+1)}{(\ell+1)-1}p_{n-1-i+(\ell+1)}\\ \text{(by the definition of } c_{i,\ell+1})}}y_j^{(\ell+1)-1}\\ &\qquad(\text{here, we have substituted } \ell+1 \text{ for } k \text{ in the sum})\\ &=\sum_{\ell=0}^{n-1}\underbrace{\dbinom{n-1-i+(\ell+1)}{(\ell+1)-1}}_{=\dbinom{n-i+\ell}{\ell}}\underbrace{p_{n-1-i+(\ell+1)}}_{=p_{n-i+\ell}}\underbrace{y_j^{(\ell+1)-1}}_{=y_j^{\ell}}\\ &=\sum_{\ell=0}^{n-1}\dbinom{n-i+\ell}{\ell}p_{n-i+\ell}y_j^{\ell}.\end{aligned}$$

Hence, (1136) rewrites as

$$CD=\left(\sum_{\ell=0}^{n-1}\dbinom{n-i+\ell}{\ell}p_{n-i+\ell}y_j^{\ell}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Now, $B=\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $CD=\left(\sum_{\ell=0}^{n-1}\dbinom{n-i+\ell}{\ell}p_{n-i+\ell}y_j^{\ell}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence, the definition of the product of two matrices shows that

$$B(CD)=\left(\sum_{k=1}^{n}x_i^{n-k}\left(\sum_{\ell=0}^{n-1}\dbinom{n-k+\ell}{\ell}p_{n-k+\ell}y_j^{\ell}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}. \tag{1137}$$

Now, let $(i,j) \in \{1,2,\ldots,n\}^2$. Then,

$$\sum_{k=1}^{n} x_i^{n-k} \left(\sum_{\ell=0}^{n-1} \binom{n-k+\ell}{\ell} p_{n-k+\ell} y_j^{\ell} \right)$$
$$= \sum_{k=0}^{n-1} x_i^{k} \left(\sum_{\ell=0}^{n-1} \binom{k+\ell}{\ell} p_{k+\ell} y_j^{\ell} \right)$$
(here, we have substituted k for $n-k$ in the first sum)
$$= \sum_{k=0}^{n-1} \sum_{\ell=0}^{n-1} x_i^{k} \binom{k+\ell}{\ell} p_{k+\ell} y_j^{\ell}$$
$$= \sum_{k=0}^{n-1} \sum_{\ell=k}^{n-1+k} x_i^{k} \underbrace{\binom{k+(\ell-k)}{\ell-k}}_{=\binom{\ell}{\ell-k}} \underbrace{p_{k+(\ell-k)}}_{=p_\ell} y_j^{\ell-k}$$
(here, we have substituted $\ell-k$ for ℓ in the second sum)
$$= \sum_{k=0}^{n-1} \underbrace{\sum_{\ell=k}^{n-1+k} x_i^{k} \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k}}_{\substack{=\sum_{\ell=k}^{n-1} x_i^k \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k} + \sum_{\ell=n}^{n-1+k} x_i^k \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k} \\ \text{(since } k\leq n\leq n-1+k+1 \text{ (since } n\leq n+k=n-1+k+1))}}$$
$$= \sum_{k=0}^{n-1} \left(\sum_{\ell=k}^{n-1} x_i^{k} \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k} + \sum_{\ell=n}^{n-1+k} x_i^{k} \binom{\ell}{\ell-k} \underbrace{p_\ell}_{\substack{=0 \\ \text{(by (1132)} \\ \text{(since } \ell\in\{n,n+1,\ldots,n-1+k\} \\ \subseteq\{n,n+1,n+2,\ldots\}))}} y_j^{\ell-k} \right)$$
$$= \sum_{k=0}^{n-1} \left(\sum_{\ell=k}^{n-1} x_i^{k} \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k} + \underbrace{\sum_{\ell=n}^{n-1+k} x_i^{k} \binom{\ell}{\ell-k} 0 y_j^{\ell-k}}_{=0} \right)$$
$$= \sum_{k=0}^{n-1} \sum_{\ell=k}^{n-1} x_i^{k} \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k}. \qquad (1138)$$

On the other hand, $P(X) = \sum_{k=0}^{n-1} p_k X^k = \sum_{\ell=0}^{n-1} p_\ell X^\ell$ (here, we renamed the summation index k as ℓ). Hence,

$$\begin{aligned}
P\left(x_i+y_j\right) &= \sum_{\ell=0}^{n-1} p_\ell \underbrace{\left(x_i+y_j\right)^\ell}_{\substack{=\sum_{k=0}^{\ell}\binom{\ell}{k}x_i^k y_j^{\ell-k} \\ \text{(by the binomial formula)}}} = \sum_{\ell=0}^{n-1} p_\ell \sum_{k=0}^{\ell}\binom{\ell}{k}x_i^k y_j^{\ell-k} \\
&= \underbrace{\sum_{\ell=0}^{n-1}\sum_{k=0}^{\ell}}_{=\sum_{\substack{(\ell,k)\in\{0,1,\ldots,n-1\}^2; \\ k\leq\ell}} = \sum_{k=0}^{n-1}\sum_{\ell=k}^{n-1}} p_\ell \binom{\ell}{k} x_i^k y_j^{\ell-k} \\
&= \sum_{k=0}^{n-1}\sum_{\ell=k}^{n-1} p_\ell \underbrace{\binom{\ell}{k}}_{\substack{=\binom{\ell}{\ell-k} \\ \text{(by (232), applied to } \ell \text{ and } k \\ \text{instead of } m \text{ and } n\text{)}}} x_i^k y_j^{\ell-k} \\
&= \sum_{k=0}^{n-1}\sum_{\ell=k}^{n-1} p_\ell \binom{\ell}{\ell-k} x_i^k y_j^{\ell-k} = \sum_{k=0}^{n-1}\sum_{\ell=k}^{n-1} x_i^k \binom{\ell}{\ell-k} p_\ell y_j^{\ell-k} \\
&= \sum_{k=1}^{n} x_i^{n-k} \left(\sum_{\ell=0}^{n-1} \binom{n-k+\ell}{\ell} p_{n-k+\ell} y_j^\ell\right) \qquad \text{(by (1138))}. \qquad (1139)
\end{aligned}$$

Now, let us forget that we fixed (i,j). We thus have proven (1139) for every $(i,j)\in\{1,2,\ldots,n\}^2$. Hence,

$$\begin{aligned}
&\left(P\left(x_i+y_j\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n} \\
&= \left(\sum_{k=1}^{n} x_i^{n-k} \left(\sum_{\ell=0}^{n-1} \binom{n-k+\ell}{\ell} p_{n-k+\ell} y_j^\ell\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n} = B\left(CD\right) \qquad \text{(by (1137))} \\
&= BCD.
\end{aligned}$$

Hence,

$$\det \underbrace{\left(\left(P\left(x_i + y_j \right) \right)_{1\leq i\leq n,\ 1\leq j\leq n} \right)}_{=BCD}$$
$$= \det\left(BCD\right) = \det B \cdot \underbrace{\det\left(CD\right)}_{\substack{=\det C \cdot \det D \\ \text{(by Theorem 6.23, applied to } C \text{ and } D \\ \text{instead of } A \text{ and } B\text{)}}}$$
$$\left(\text{by Theorem 6.23, applied to } B \text{ and } CD \text{ instead of } A \text{ and } B\right)$$
$$= \underbrace{\det B}_{\substack{= \prod\limits_{1\leq i<j\leq n} \left(x_i - x_j\right) \\ \text{(by (1133))}}} \cdot \underbrace{\det C}_{\substack{= p_{n-1}^n \left(\prod\limits_{k=0}^{n-1} \binom{n-1}{k} \right) \\ \text{(by (1134))}}} \cdot \underbrace{\det D}_{\substack{= \prod\limits_{1\leq i<j\leq n} \left(y_j - y_i\right) \\ \text{(by (1135))}}}$$
$$= \left(\prod_{1\leq i<j\leq n} \left(x_i - x_j\right) \right) \cdot p_{n-1}^n \left(\prod_{k=0}^{n-1} \binom{n-1}{k} \right) \cdot \left(\prod_{1\leq i<j\leq n} \left(y_j - y_i\right) \right)$$
$$= p_{n-1}^n \left(\prod_{k=0}^{n-1} \binom{n-1}{k} \right) \left(\prod_{1\leq i<j\leq n} \left(x_i - x_j\right) \right) \left(\prod_{1\leq i<j\leq n} \left(y_j - y_i\right) \right).$$

This solves Exercise 6.17 **(c)**.

(a) For any two objects i and j, we define $\delta_{i,j}$ to be the element $\begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$ of $\mathbb{K}$.

Let $m \in \{0, 1, \ldots, n-1\}$. (You are reading this right: We are not requiring m to belong to $\{0, 1, \ldots, n-2\}$; the purpose of this is to obtain a result that will bring us close to solving both parts **(a)** and **(b)** simultaneously.)

Define an n-tuple $\left(p_0, p_1, \ldots, p_{n-1}\right) \in \mathbb{K}^n$ by

$$\left(p_k = \delta_{k,m} \qquad \text{for every } k \in \{0, 1, \ldots, n-1\} \right).$$

Let $P\left(X\right) \in \mathbb{K}\left[X\right]$ be the polynomial $\sum\limits_{k=0}^{n-1} p_k X^k$.

We have

$$\begin{aligned} P(X) &= \underbrace{\sum_{k=0}^{n-1}}_{=\sum_{k\in\{0,1,\ldots,n-1\}}} \underbrace{p_k}_{=\delta_{k,m}} X^k = \sum_{k\in\{0,1,\ldots,n-1\}} \delta_{k,m} X^k \\ &= \underbrace{\delta_{m,m}}_{\substack{=1\\ (\text{since } m=m)}} X^m + \sum_{\substack{k\in\{0,1,\ldots,n-1\};\\ k\neq m}} \underbrace{\delta_{k,m}}_{\substack{=0\\ (\text{since } k\neq m)}} X^k \\ &\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } k=m \text{ from the sum}\\ (\text{since } m\in\{0,1,\ldots,n-1\})\end{array}\right) \\ &= X^m + \underbrace{\sum_{\substack{k\in\{0,1,\ldots,n-1\};\\ k\neq m}} 0X^k}_{=0} = X^m. \end{aligned}$$

Every $(i,j)\in\{1,2,\ldots,n\}^2$ satisfies $P(x_i+y_j) = (x_i+y_j)^m$ (since $P(X)=X^m$). In other words, $(P(x_i+y_j))_{1\leq i\leq n,\ 1\leq j\leq n} = \left((x_i+y_j)^m\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence,

$$\left((x_i+y_j)^m\right)_{1\leq i\leq n,\ 1\leq j\leq n} = (P(x_i+y_j))_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Taking determinants on both sides of this equality, we obtain

$$\begin{aligned} &\det\left(\left((x_i+y_j)^m\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ &= \det\left((P(x_i+y_j))_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ &= \underbrace{p_{n-1}^n}_{\substack{=\delta_{n-1,m}^n \\ (\text{since } p_{n-1}=\delta_{n-1,m} \\ (\text{by the definition of } p_{n-1}))}} \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}(x_i-x_j)\right)\left(\prod_{1\leq i<j\leq n}(y_j-y_i)\right) \\ &\qquad (\text{by Exercise 6.17 } \textbf{(c)}) \\ &= \delta_{n-1,m}^n \left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}(x_i-x_j)\right)\left(\prod_{1\leq i<j\leq n}(y_j-y_i)\right). \end{aligned} \tag{1140}$$

Now, let us forget that we fixed m. We thus have proven (1140) for every $m\in\{0,1,\ldots,n-1\}$.

Now, let $m\in\{0,1,\ldots,n-2\}$. Thus, $m\neq n-1$, so that $\delta_{n-1,m}=0$. Hence, $\delta_{n-1,m}^n = 0^n = 0$ (since n is a positive integer).

On the other hand, $m\in\{0,1,\ldots,n-2\}\subseteq\{0,1,\ldots,n-1\}$, and therefore (1140) holds. Hence,

$$\begin{aligned}
&\det\left(\left(\left(x_i+y_j\right)^m\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\
&=\underbrace{\delta_{n-1,m}^n}_{=0}\left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}\left(x_i-x_j\right)\right)\left(\prod_{1\leq i<j\leq n}\left(y_j-y_i\right)\right) \\
&=0.
\end{aligned}$$

This solves Exercise 6.17 **(a)**.

(b) For any two objects i and j, we define $\delta_{i,j}$ as in the solution to Exercise 6.17 **(a)**.

We have $\delta_{n-1,n-1}=1$ and thus $\delta_{n-1,n-1}^n=1^n=1$.

In our solution of Exercise 6.17 **(a)**, we have proven the equality (1140) for every $m\in\{0,1,\ldots,n-1\}$. Thus, we can apply this equality to $m=n-1$. As a result, we obtain

$$\begin{aligned}
&\det\left(\left(\left(x_i+y_j\right)^{n-1}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\
&=\underbrace{\delta_{n-1,n-1}^n}_{=1}\left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}\left(x_i-x_j\right)\right)\left(\prod_{1\leq i<j\leq n}\left(y_j-y_i\right)\right) \\
&=\left(\prod_{k=0}^{n-1}\binom{n-1}{k}\right)\left(\prod_{1\leq i<j\leq n}\left(x_i-x_j\right)\right)\left(\prod_{1\leq i<j\leq n}\left(y_j-y_i\right)\right).
\end{aligned}$$

This solves Exercise 6.17 **(b)**. □

7.85. Solution to Exercise 6.18

We shall give a detailed solution to Exercise 6.18 in Section 7.86 further below. For now, let us give some pointers to the literature.

Exercise 6.18 is obtained from [GriRei18, Exercise 2.7.8(a)] by setting $a_i=x_i$ and $b_j=-y_j$. (See the ancillary PDF file of the arXiv version of [GriRei18] for the solutions to the exercises.)

Exercise 6.18 can also be obtained from [Grinbe09, Theorem 2] by setting $k=\mathbb{K}$,

$m = n$, $a_j = y_j$ and $b_i = -x_i$ (and observing that

$$\prod_{\substack{(i,j)\in\{1,2,\ldots,n\}^2;\\ i>j}} \underbrace{\left(\left(y_i - y_j\right)\left(\left(-x_j\right) - \left(-x_i\right)\right)\right)}_{\substack{=\left(y_j-y_i\right)\left(x_j-x_i\right)\\ =\left(x_j-x_i\right)\left(y_j-y_i\right)}}$$

$$= \prod_{\substack{(i,j)\in\{1,2,\ldots,n\}^2;\\ i>j}} \left(\left(x_j - x_i\right)\left(y_j - y_i\right)\right) = \prod_{\substack{(j,i)\in\{1,2,\ldots,n\}^2;\\ j>i}} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right)$$

(here, we have renamed the index (i,j) as (j,i))

$$= \prod_{1\leq i<j\leq n} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right)$$

). The statement of [Grinbe09, Theorem 2] makes the requirement that k be a field; however, this is easily seen to be unnecessary for the proof.

7.86. Solution to Exercise 6.19

7.86.1. The solution

Exercise 6.19 is the equality (12.57.3) in [GriRei18, ancillary PDF file]. The following solution is taken from [GriRei18, ancillary PDF file].

Solution to Exercise 6.19. We shall use the Iverson bracket notation (introduced in Definition 3.48).

We have $n \in \{1,2,\ldots,n\}$ (since n is a positive integer).

Define an $n\times n$-matrix $A \in \mathbb{K}^{n\times n}$ by $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

For every $(i,j) \in \{1,2,\ldots,n\}^2$, define an element $b_{i,j}$ of $\mathbb{K}$ by

$$b_{i,j} = [i=j]\, a_{n,n} - [i=n \text{ and } j\neq n]\, a_{n,j}. \tag{1141}$$

Then, it is easy to see that

$$b_{i,j} = 0 \text{ for every } (i,j) \in \{1,2,\ldots,n\}^2 \text{ satisfying } i<j. \tag{1142}$$

[*Proof of (1142):* Let $(i,j) \in \{1,2,\ldots,n\}^2$ be such that $i<j$. Then, $i\neq j$ (since $i<j$), and thus we don't have $i=j$. Hence, $[i=j]=0$. Also, $j\in\{1,2,\ldots,n\}$ (since $(i,j)\in\{1,2,\ldots,n\}^2$) and thus $j\leq n$, so that $i<j\leq n$ and thus $i\neq n$. Hence, we don't have $i=n$. Thus, we don't have $(i=n \text{ and } j\neq n)$ either. In other words, $[i=n \text{ and } j\neq n]=0$. Now, (1141) yields $b_{i,j} = \underbrace{[i=j]}_{=0} a_{n,n} - \underbrace{[i=n \text{ and } j\neq n]}_{=0} a_{n,j} = \underbrace{0a_{n,n}}_{=0} - \underbrace{0a_{n,j}}_{=0} = 0$. This proves (1142).]

Define an $n\times n$-matrix $B\in\mathbb{K}^{n\times n}$ by $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Thus, Exercise 6.3 (applied to B and $b_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det B = b_{1,1}b_{2,2}\cdots b_{n,n} \tag{1143}$$

(since $b_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$).

But

$$\text{every } i \in \{1,2,\ldots,n\} \text{ satisfies } b_{i,i} = a_{n,n}. \qquad (1144)$$

[*Proof of (1144):* Let $i \in \{1,2,\ldots,n\}$. Then, we don't have $(i = n \text{ and } i \neq n)$ (because the two statements $i = n$ and $i \neq n$ clearly contradict each other). In other words, $[i = n \text{ and } i \neq n] = 0$. Now, (1141) (applied to $j = i$) yields

$$b_{i,i} = \underbrace{[i = i]}_{\substack{=1 \\ \text{(since } i=i \text{ is true)}}} a_{n,n} - \underbrace{[i = n \text{ and } i \neq n]}_{=0} a_{n,i} = \underbrace{1a_{n,n}}_{=a_{n,n}} - \underbrace{0a_{n,i}}_{=0} = a_{n,n}.$$

This proves (1144).]

Thus, (1143) becomes

$$\det B = b_{1,1}b_{2,2}\cdots b_{n,n} = \prod_{i=1}^{n} \underbrace{b_{i,i}}_{\substack{=a_{n,n} \\ \text{(by (1144))}}} = \prod_{i=1}^{n} a_{n,n} = a_{n,n}^{n}. \qquad (1145)$$

For each $(i,j) \in \{1,2,\ldots,n\}^2$, we define an element $c_{i,j}$ of $\mathbb{K}$ by

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j}. \qquad (1146)$$

But we have $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. The definition of AB thus yields

$$AB = \left(\underbrace{\sum_{k=1}^{n} a_{i,k}b_{k,j}}_{\substack{=c_{i,j} \\ \text{(by (1146))}}} \right)_{1\leq i\leq n,\ 1\leq j\leq n} = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

But Theorem 6.23 yields

$$\det(AB) = \det A \cdot \underbrace{\det B}_{\substack{=a_{n,n}^{n} \\ \text{(by (1145))}}} = (\det A)\cdot a_{n,n}^{n} = a_{n,n}^{n} \det A. \qquad (1147)$$

But if i and j are any two elements of $\{1,2,\ldots,n\}$, then

$$c_{i,j} = a_{i,j}a_{n,n} - [j \neq n]\, a_{i,n}a_{n,j}. \qquad (1148)$$

[*Proof of (1148):* Let i and j be two elements of $\{1,2,\ldots,n\}$. The equality (1146) yields

$$\begin{aligned}
c_{i,j} &= \underbrace{\sum_{k=1}^{n}}_{=\sum_{k\in\{1,2,\ldots,n\}}} a_{i,k} \underbrace{b_{k,j}}_{\substack{=[k=j]a_{n,n}-[k=n \text{ and } j\neq n]a_{n,j} \\ \text{(by (1141) (applied to } k \text{ instead of } i\text{))}}} \\
&= \sum_{k\in\{1,2,\ldots,n\}} \underbrace{a_{i,k}\left([k=j]\,a_{n,n}-[k=n \text{ and } j\neq n]\,a_{n,j}\right)}_{=[k=j]a_{i,k}a_{n,n}-[k=n \text{ and } j\neq n]a_{i,k}a_{n,j}} \\
&= \sum_{k\in\{1,2,\ldots,n\}} \left([k=j]\,a_{i,k}a_{n,n}-[k=n \text{ and } j\neq n]\,a_{i,k}a_{n,j}\right) \\
&= \sum_{k\in\{1,2,\ldots,n\}} [k=j]\,a_{i,k}a_{n,n} - \sum_{k\in\{1,2,\ldots,n\}} [k=n \text{ and } j\neq n]\,a_{i,k}a_{n,j}. \qquad (1149)
\end{aligned}$$

But

$$\begin{aligned}
&\sum_{k\in\{1,2,\ldots,n\}} [k=j]\,a_{i,k}a_{n,n} \\
&= \underbrace{[j=j]}_{\substack{=1 \\ \text{(since } j=j \text{ holds)}}} a_{i,j}a_{n,n} + \sum_{\substack{k\in\{1,2,\ldots,n\}; \\ k\neq j}} \underbrace{[k=j]}_{\substack{=0 \\ \text{(since } k=j \text{ does not hold} \\ \text{(since } k\neq j\text{))}}} a_{i,k}a_{n,n} \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } k=j \text{ from the sum} \\ \text{(since } j\in\{1,2,\ldots,n\}\text{)}\end{array}\right) \\
&= a_{i,j}a_{n,n} + \underbrace{\sum_{\substack{k\in\{1,2,\ldots,n\}; \\ k\neq j}} 0a_{i,k}a_{n,n}}_{=0} = a_{i,j}a_{n,n}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{k\in\{1,2,\ldots,n\}} [k=n \text{ and } j\neq n]\,a_{i,k}a_{n,j} \\
&= \underbrace{[n=n \text{ and } j\neq n]}_{\substack{=[j\neq n] \\ \text{(since the statement } (n=n \text{ and } j\neq n) \\ \text{is equivalent to } (j\neq n) \\ \text{(since } n=n \text{ always holds))}}} a_{i,n}a_{n,j} + \sum_{\substack{k\in\{1,2,\ldots,n\}; \\ k\neq n}} \underbrace{[k=n \text{ and } j\neq n]}_{\substack{=0 \\ \text{(since } (k=n \text{ and } j\neq n) \text{ does not hold} \\ \text{(since } k=n \text{ does not hold} \\ \text{(since } k\neq n\text{)))}}} a_{i,k}a_{n,j} \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } k=n \text{ from the sum} \\ \text{(since } n\in\{1,2,\ldots,n\}\text{)}\end{array}\right) \\
&= [j\neq n]\,a_{i,n}a_{n,j} + \underbrace{\sum_{\substack{k\in\{1,2,\ldots,n\}; \\ k\neq n}} 0a_{i,k}a_{n,j}}_{=0} = [j\neq n]\,a_{i,n}a_{n,j}.
\end{aligned}$$

Hence, (1149) becomes

$$c_{i,j} = \underbrace{\sum_{k\in\{1,2,\ldots,n\}} [k=j] \, a_{i,k} a_{n,n}}_{= a_{i,j} a_{n,n}} - \underbrace{\sum_{k\in\{1,2,\ldots,n\}} [k=n \text{ and } j \neq n] \, a_{i,k} a_{n,j}}_{= [j\neq n] a_{i,n} a_{n,j}}$$
$$= a_{i,j} a_{n,n} - [j \neq n] \, a_{i,n} a_{n,j}.$$

This proves (1148).]

Now, we can easily check that

$$c_{n,j} = 0 \text{ for every } j \in \{1, 2, \ldots, n-1\}. \tag{1150}$$

[*Proof of (1150):* Let $j \in \{1, 2, \ldots, n-1\}$. Then, $j \leq n-1 < n$, so that $j \neq n$. Thus, $[j \neq n] = 1$. Also, $j \in \{1, 2, \ldots, n-1\} \subseteq \{1, 2, \ldots, n\}$ and $n \in \{1, 2, \ldots, n\}$. Hence, n and j are two elements of $\{1, 2, \ldots, n\}$. Thus, (1148) (applied to $i = n$) yields

$$c_{n,j} = \underbrace{a_{n,j} a_{n,n}}_{= a_{n,n} a_{n,j}} - \underbrace{[j \neq n]}_{=1} a_{n,n} a_{n,j} = a_{n,n} a_{n,j} - a_{n,n} a_{n,j} = 0.$$

This proves (1150).]

So we have shown that $c_{n,j} = 0$ for every $j \in \{1, 2, \ldots, n-1\}$. Hence, Theorem 6.43 (applied to AB and $c_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det(AB) = c_{n,n} \cdot \det\left(\left(c_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right)$$

(since $AB = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Comparing this with (1147), we find

$$a_{n,n}^n \det A = c_{n,n} \cdot \det\left(\left(c_{i,j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq n-1}\right). \tag{1151}$$

But every $(i,j) \in \{1, 2, \ldots, n-1\}^2$ satisfies

$$c_{i,j} = a_{i,j} a_{n,n} - a_{i,n} a_{n,j}. \tag{1152}$$

[*Proof of (1152):* Let $(i,j) \in \{1, 2, \ldots, n-1\}^2$. Since $(i,j) \in \{1, 2, \ldots, n-1\}^2$, we have $i \in \{1, 2, \ldots, n-1\}$ and $j \in \{1, 2, \ldots, n-1\}$. Since $j \in \{1, 2, \ldots, n-1\}$, we have $j \leq n-1 < n$, so that $j \neq n$. Thus, $[j \neq n] = 1$.

From $i \in \{1, 2, \ldots, n-1\} \subseteq \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n-1\} \subseteq \{1, 2, \ldots, n\}$, we conclude that i and j are two elements of $\{1, 2, \ldots, n\}$. Hence, (1148) yields $c_{i,j} = a_{i,j} a_{n,n} - \underbrace{[j \neq n]}_{=1} a_{i,n} a_{n,j} = a_{i,j} a_{n,n} - a_{i,n} a_{n,j}$. This proves (1152).]

Furthermore,

$$c_{n,n} = a_{n,n}^2. \tag{1153}$$

[*Proof of (1153):* We have $n \in \{1, 2, \ldots, n\}$. Thus, n and n are two elements of $\{1, 2, \ldots, n\}$. Hence, (1148) (applied to $i = n$ and $j = n$) yields

$$c_{n,n} = a_{n,n} a_{n,n} - \underbrace{[n \neq n]}_{\substack{=0 \\ \text{(since we don't have } n \neq n)}} a_{n,n} a_{n,n} = a_{n,n} a_{n,n} - \underbrace{0 a_{n,n} a_{n,n}}_{=0} = a_{n,n} a_{n,n} = a_{n,n}^2.$$

This proves (1153).]

Now, (1151) becomes

$$\begin{aligned} a_{n,n}^n \det A &= \underbrace{c_{n,n}}_{\substack{=a_{n,n}^2 \\ \text{(by (1153))}}} \cdot \det \left(\left(\underbrace{c_{i,j}}_{\substack{=a_{i,j}a_{n,n} - a_{i,n}a_{n,j} \\ \text{(by (1152))}}} \right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \right) \\ &= a_{n,n}^2 \cdot \det \left(\left(a_{i,j} a_{n,n} - a_{i,n} a_{n,j} \right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \right). \end{aligned}$$

We can divide both sides of this equality by $a_{n,n}^2$ (since $a_{n,n}^2$ is invertible in $\mathbb{K}$ (because $a_{n,n}$ is invertible in $\mathbb{K}$)), and thus obtain

$$\frac{1}{a_{n,n}^2} a_{n,n}^n \det A = \det \left(\left(a_{i,j} a_{n,n} - a_{i,n} a_{n,j} \right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \right).$$

Thus,

$$\begin{aligned} \det \left(\left(a_{i,j} a_{n,n} - a_{i,n} a_{n,j} \right)_{1 \leq i \leq n-1,\ 1 \leq j \leq n-1} \right) &= \underbrace{\frac{1}{a_{n,n}^2} a_{n,n}^n}_{=a_{n,n}^{n-2}} \det \underbrace{A}_{=\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}} \\ &= a_{n,n}^{n-2} \cdot \det \left(\left(a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right). \end{aligned}$$

This solves Exercise 6.19. □

7.86.2. Solution to Exercise 6.18

We shall now observe our promise and solve Exercise 6.18 using Exercise 6.19. Again, our solution will follow [GriRei18, ancillary PDF file].

We begin with some elementary lemmas:

Lemma 7.164. Let $n \in \mathbb{N}$. Let $\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix. Let $c_1, c_2, \ldots, c_n$ be n elements of $\mathbb{K}$. Then,

$$\det \left(\left(c_i a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) = \left(\prod_{i=1}^{n} c_i \right) \cdot \det \left(\left(a_{i,j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right).$$

Proof of Lemma 7.164. Define an $n \times n$-matrix $A \in \mathbb{K}^{n \times n}$ by $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Define an $n \times n$-matrix $C \in \mathbb{K}^{n \times n}$ by $C = \left(c_i a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, (341) (applied to C and $c_i a_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det C = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^n \left(c_i a_{i,\sigma(i)}\right)}_{= \left(\prod_{i=1}^n c_i\right)\left(\prod_{i=1}^n a_{i,\sigma(i)}\right)} = \sum_{\sigma \in S_n} (-1)^\sigma \left(\prod_{i=1}^n c_i\right)\left(\prod_{i=1}^n a_{i,\sigma(i)}\right)$$
$$= \left(\prod_{i=1}^n c_i\right) \cdot \underbrace{\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}}_{\substack{= \det A \\ \text{(by (341))}}} = \left(\prod_{i=1}^n c_i\right) \cdot \det \underbrace{A}_{= \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}}$$
$$= \left(\prod_{i=1}^n c_i\right) \cdot \det \left(\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right).$$

In view of $C = \left(c_i a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, this rewrites as

$$\det \left(\left(c_i a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = \left(\prod_{i=1}^n c_i\right) \cdot \det \left(\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right).$$

This proves Lemma 7.164. □

Lemma 7.165. Let $n \in \mathbb{N}$. Let $\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix. Let $d_1, d_2, \ldots, d_n$ be n elements of $\mathbb{K}$. Then,

$$\det \left(\left(d_j a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = \left(\prod_{j=1}^n d_j\right) \cdot \det \left(\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right).$$

Proof of Lemma 7.165. Let $\sigma \in S_n$. Thus, σ is a permutation of the set $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$). In other words, σ is a bijection $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. Now,

$$\underbrace{\prod_{i=1}^n}_{= \prod_{i \in \{1,2,\ldots,n\}}} d_{\sigma(i)} = \prod_{i \in \{1,2,\ldots,n\}} d_{\sigma(i)} = \underbrace{\prod_{j \in \{1,2,\ldots,n\}}}_{= \prod_{j=1}^n} d_j$$
$$\left(\begin{array}{c}\text{here, we have substituted } j \text{ for } \sigma(i) \text{ in the product,} \\ \text{since } \sigma \text{ is a bijection } \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}\end{array}\right)$$
$$= \prod_{j=1}^n d_j. \tag{1154}$$

Now, forget that we fixed σ. We thus have proven the equality (1154) for each $\sigma \in S_n$.

Define an $n \times n$-matrix $A \in \mathbb{K}^{n \times n}$ by $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Define an $n \times n$-matrix $D \in \mathbb{K}^{n \times n}$ by $D = \left(d_j a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, (341) (applied to D and $d_j a_{i,j}$ instead of A and $a_{i,j}$) yields

$$\det D = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\prod_{i=1}^{n} \left(d_{\sigma(i)} a_{i,\sigma(i)}\right)}_{= \left(\prod_{i=1}^{n} d_{\sigma(i)}\right) \left(\prod_{i=1}^{n} a_{i,\sigma(i)}\right)} = \sum_{\sigma \in S_n} (-1)^\sigma \underbrace{\left(\prod_{i=1}^{n} d_{\sigma(i)}\right)}_{\substack{= \prod_{j=1}^{n} d_j \\ \text{(by (1154))}}} \left(\prod_{i=1}^{n} a_{i,\sigma(i)}\right)$$

$$= \sum_{\sigma \in S_n} (-1)^\sigma \left(\prod_{j=1}^{n} d_j\right) \left(\prod_{i=1}^{n} a_{i,\sigma(i)}\right) = \left(\prod_{j=1}^{n} d_j\right) \cdot \underbrace{\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{= \det A \\ \text{(by (341))}}}$$

$$= \left(\prod_{j=1}^{n} d_j\right) \cdot \det \underbrace{A}_{= \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}} = \left(\prod_{j=1}^{n} d_j\right) \cdot \det \left(\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right).$$

In view of $D = \left(d_j a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, this rewrites as

$$\det \left(\left(d_j a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) = \left(\prod_{j=1}^{n} d_j\right) \cdot \det \left(\left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right).$$

This proves Lemma 7.165. □

Next, we will show a proposition that generalizes Exercise 6.18:

Proposition 7.166. Let $n \in \mathbb{N}$. For every $i \in \{1, 2, \ldots, n\}$, let a_i, b_i, c_i and d_i be four elements of $\mathbb{K}$. Assume that $a_i d_j - b_i c_j$ is an invertible element of $\mathbb{K}$ for every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$. Then,

$$\det \left(\left(\frac{1}{a_i d_j - b_i c_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) = \frac{\prod_{1 \leq j < i \leq n} \left(\left(a_i b_j - a_j b_i\right) \left(c_j d_i - c_i d_j\right) \right)}{\prod_{(i,j) \in \{1,2,\ldots,n\}^2} \left(a_i d_j - b_i c_j\right)}$$

Proof of Proposition 7.166. For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$, we define an element $x_{i,j}$ of $\mathbb{K}$ by

$$x_{i,j} = a_i d_j - b_i c_j. \tag{1155}$$

We have assumed that $a_i d_j - b_i c_j$ is an invertible element of $\mathbb{K}$ for every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$. In other words, $x_{i,j}$ is an invertible element of $\mathbb{K}$

for every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ (because $x_{i,j} = a_i d_j - b_i c_j$ for every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$).

We are going to show that

$$\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq k,\ 1\leq j\leq k}\right) = \frac{\prod\limits_{1\leq j<i\leq k}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)}{\prod\limits_{(i,j)\in\{1,2,\ldots,k\}^2} x_{i,j}} \tag{1156}$$

for every $k \in \{0, 1, \ldots, n\}$.

[*Proof of (1156):* We will prove (1156) by induction over k:

Induction base: It is easy to see that (1156) holds if $k = 0$ [514]. Hence, the induction base is complete.

Induction step: Let $m \in \{1, 2, \ldots, n\}$. Assume that (1156) is proven for $k = m - 1$. We need to show that (1156) holds for $k = m$.

---

[514] *Proof.* Assume that $k = 0$. Then, $\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq k,\ 1\leq j\leq k} = \left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq 0,\ 1\leq j\leq 0}$ is a 0×0-matrix, and thus has determinant 1 (since every 0×0-matrix has determinant 1). In other words, $\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq k,\ 1\leq j\leq k}\right) = 1$.

But then there exist no integers i and j satisfying $1 \leq j < i \leq k$ (because any such integers would have to satisfy $1 \leq j < i \leq k = 0 < 1$, which is absurd). Thus, $\prod\limits_{1\leq j<i\leq k}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)$ is an empty product. Hence, $\prod\limits_{1\leq j<i\leq k}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right) = (\text{empty product}) = 1$. Also, since $k = 0$, we have $\{1, 2, \ldots, k\} = \{1, 2, \ldots, 0\} = \varnothing$, so that $\{1, 2, \ldots, k\}^2 = \varnothing^2 = \varnothing$. Thus, $\prod\limits_{(i,j)\in\{1,2,\ldots,k\}^2} x_{i,j}$ is an empty product. Hence, $\prod\limits_{(i,j)\in\{1,2,\ldots,k\}^2} x_{i,j} = (\text{empty product}) = 1$.

Now, dividing the equality $\prod\limits_{1\leq j<i\leq k}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right) = 1$ by the equality $\prod\limits_{(i,j)\in\{1,2,\ldots,k\}^2} x_{i,j} = 1$, we obtain

$$\frac{\prod\limits_{1\leq j<i\leq k}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)}{\prod\limits_{(i,j)\in\{1,2,\ldots,k\}^2} x_{i,j}} = \frac{1}{1} = 1.$$

Comparing this with $\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq k,\ 1\leq j\leq k}\right) = 1$, we obtain $\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq k,\ 1\leq j\leq k}\right) = \frac{\prod\limits_{1\leq j<i\leq k}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)}{\prod\limits_{(i,j)\in\{1,2,\ldots,k\}^2} x_{i,j}}$. Thus, (1156) holds if $k = 0$, qed.

We know that (1156) is proven for $k = m-1$. Thus,

$$\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right) = \frac{\prod\limits_{1\leq j<i\leq m-1}\left(\left(a_ib_j - a_jb_i\right)\left(c_jd_i - c_id_j\right)\right)}{\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}}. \qquad (1157)$$

We recall that $x_{i,j}$ is an invertible element of $\mathbb{K}$ for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$. Applying this to $i = m$ and $j = m$, we conclude that $x_{m,m}$ is an invertible element of $\mathbb{K}$. Hence, $\frac{1}{x_{m,m}}$ is an invertible element of $\mathbb{K}$ as well (because it is the inverse of $x_{m,m}$). Thus, Exercise 6.19 (applied to m and $\frac{1}{x_{i,j}}$ instead of n and $a_{i,j}$) yields

$$\begin{aligned}
&\det\left(\left(\frac{1}{x_{i,j}}\cdot\frac{1}{x_{m,m}} - \frac{1}{x_{i,m}}\cdot\frac{1}{x_{m,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right) \\
&= \left(\frac{1}{x_{m,m}}\right)^{m-2}\cdot\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m,\ 1\leq j\leq m}\right).
\end{aligned}$$

Multiplying both sides of this equality with $x_{m,m}^{m-2}$, we obtain

$$\begin{aligned}
&x_{m,m}^{m-2}\cdot\det\left(\left(\frac{1}{x_{i,j}}\cdot\frac{1}{x_{m,m}} - \frac{1}{x_{i,m}}\cdot\frac{1}{x_{m,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right) \\
&= x_{m,m}^{m-2}\cdot\underbrace{\left(\frac{1}{x_{m,m}}\right)^{m-2}}_{=\frac{1}{x_{m,m}^{m-2}}}\cdot\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m,\ 1\leq j\leq m}\right) \\
&= \underbrace{x_{m,m}^{m-2}\cdot\frac{1}{x_{m,m}^{m-2}}}_{=1}\cdot\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m,\ 1\leq j\leq m}\right) \\
&= \det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m,\ 1\leq j\leq m}\right). \qquad (1158)
\end{aligned}$$

But every $(i,j) \in \{1,2,\ldots,m-1\}^2$ satisfies

$$\begin{aligned}
&\underbrace{x_{i,m}}_{\substack{=a_i d_m - b_i c_m \\ \text{(by the definition of } x_{i,m})}} \underbrace{x_{m,j}}_{\substack{=a_m d_j - b_m c_j \\ \text{(by the definition of } x_{m,j})}} - \underbrace{x_{i,j}}_{\substack{=a_i d_j - b_i c_j \\ \text{(by the definition of } x_{i,j})}} \underbrace{x_{m,m}}_{\substack{=a_m d_m - b_m c_m \\ \text{(by the definition of } x_{m,m})}} \\
&= \underbrace{\left(a_i d_m - b_i c_m\right)\left(a_m d_j - b_m c_j\right)}_{=a_i d_m a_m d_j - a_i d_m b_m c_j - b_i c_m a_m d_j + b_i c_m b_m c_j} - \underbrace{\left(a_i d_j - b_i c_j\right)\left(a_m d_m - b_m c_m\right)}_{=a_i d_j a_m d_m - a_i d_j b_m c_m - b_i c_j a_m d_m + b_i c_j b_m c_m} \\
&= \left(\underbrace{a_i d_m a_m d_j}_{=a_i d_j a_m d_m} - \underbrace{a_i d_m b_m c_j}_{=a_i b_m c_j d_m} - \underbrace{b_i c_m a_m d_j}_{=a_m b_i c_m d_j} + \underbrace{b_i c_m b_m c_j}_{=b_i c_j b_m c_m} \right) \\
&\qquad - \left(a_i d_j a_m d_m - \underbrace{a_i d_j b_m c_m}_{=a_i b_m c_m d_j} - \underbrace{b_i c_j a_m d_m}_{=a_m b_i c_j d_m} + b_i c_j b_m c_m \right) \\
&= \left(a_i d_j a_m d_m - a_i b_m c_j d_m - a_m b_i c_m d_j + b_i c_j b_m c_m\right) \\
&\qquad - \left(a_i d_j a_m d_m - a_i b_m c_m d_j - a_m b_i c_j d_m + b_i c_j b_m c_m\right) \\
&= a_m b_i c_j d_m - a_m b_i c_m d_j - a_i b_m c_j d_m + a_i b_m c_m d_j \\
&= \left(a_m b_i - a_i b_m\right)\left(c_j d_m - c_m d_j\right) .
\end{aligned} \tag{1159}$$

Hence, every $(i,j) \in \{1,2,\ldots,m-1\}^2$ satisfies

$$\begin{aligned}
&\underbrace{\frac{1}{x_{i,j}} \cdot \frac{1}{x_{m,m}}}_{=\frac{1}{x_{i,j} x_{m,m}}} - \underbrace{\frac{1}{x_{i,m}} \cdot \frac{1}{x_{m,j}}}_{=\frac{1}{x_{i,m} x_{m,j}}} \\
&= \frac{1}{x_{i,j} x_{m,m}} - \frac{1}{x_{i,m} x_{m,j}} = \frac{x_{i,m} x_{m,j} - x_{i,j} x_{m,m}}{x_{i,j} x_{m,m} x_{i,m} x_{m,j}} \\
&= \frac{1}{x_{i,j} x_{m,m} x_{i,m} x_{m,j}} \cdot \underbrace{\left(x_{i,m} x_{m,j} - x_{i,j} x_{m,m}\right)}_{\substack{=\left(a_m b_i - a_i b_m\right)\left(c_j d_m - c_m d_j\right) \\ \text{(by (1159))}}} \\
&= \frac{1}{x_{i,j} x_{m,m} x_{i,m} x_{m,j}} \cdot \left(a_m b_i - a_i b_m\right)\left(c_j d_m - c_m d_j\right) \\
&= \frac{c_j d_m - c_m d_j}{x_{m,m} x_{m,j}} \cdot \frac{a_m b_i - a_i b_m}{x_{i,m}} \cdot \frac{1}{x_{i,j}}.
\end{aligned}$$

Hence,

$$\det\left(\left(\underbrace{\frac{1}{x_{i,j}}\cdot\frac{1}{x_{m,m}}-\frac{1}{x_{i,m}}\cdot\frac{1}{x_{m,j}}}_{=\frac{c_jd_m-c_md_j}{x_{m,m}x_{m,j}}\cdot\frac{a_mb_i-a_ib_m}{x_{i,m}}\cdot\frac{1}{x_{i,j}}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right)$$

$$=\det\left(\left(\frac{c_jd_m-c_md_j}{x_{m,m}x_{m,j}}\cdot\frac{a_mb_i-a_ib_m}{x_{i,m}}\cdot\frac{1}{x_{i,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right)$$

$$=\left(\prod_{j=1}^{m-1}\underbrace{\frac{c_jd_m-c_md_j}{x_{m,m}x_{m,j}}}_{=\frac{1}{x_{m,m}}\cdot\frac{c_jd_m-c_md_j}{x_{m,j}}}\right)\cdot\underbrace{\det\left(\left(\frac{a_mb_i-a_ib_m}{x_{i,m}}\cdot\frac{1}{x_{i,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right)}_{\substack{=\left(\prod_{i=1}^{m-1}\frac{a_mb_i-a_ib_m}{x_{i,m}}\right)\cdot\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right)\\ \text{(by Lemma 7.164 (applied to } m-1,\ \frac{1}{x_{i,j}}\\ \text{and } \frac{a_mb_i-a_ib_m}{x_{i,m}} \text{ instead of } n,\ a_{i,j} \text{ and } c_i))}}$$

$$\left(\begin{array}{c}\text{by Lemma 7.165 (applied to } m-1,\ \frac{a_mb_i-a_ib_m}{x_{i,m}}\cdot\frac{1}{x_{i,j}} \text{ and}\\ \frac{c_jd_m-c_md_j}{x_{m,m}x_{m,j}} \text{ instead of } n,\ a_{i,j} \text{ and } d_j)\end{array}\right)$$

$$=\underbrace{\left(\prod_{j=1}^{m-1}\left(\frac{1}{x_{m,m}}\cdot\frac{c_jd_m-c_md_j}{x_{m,j}}\right)\right)}_{=\left(\prod_{j=1}^{m-1}\frac{1}{x_{m,m}}\right)\cdot\left(\prod_{j=1}^{m-1}\frac{c_jd_m-c_md_j}{x_{m,j}}\right)}\cdot\left(\prod_{i=1}^{m-1}\frac{a_mb_i-a_ib_m}{x_{i,m}}\right)$$

$$\cdot\underbrace{\det\left(\left(\frac{1}{x_{i,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right)}_{\substack{=\frac{\prod_{1\leq j<i\leq m-1}\left(\left(a_ib_j-a_jb_i\right)\left(c_jd_i-c_id_j\right)\right)}{\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}x_{i,j}}\\ \text{(by (1157))}}}$$

$$= \underbrace{\left(\prod_{j=1}^{m-1} \frac{1}{x_{m,m}}\right)}_{=\left(\frac{1}{x_{m,m}}\right)^{m-1}} \cdot \left(\prod_{j=1}^{m-1} \underbrace{\frac{c_j d_m - c_m d_j}{x_{m,j}}}_{=\frac{1}{x_{m,j}}\cdot\left(c_j d_m - c_m d_j\right)}\right) \cdot \left(\prod_{i=1}^{m-1} \underbrace{\frac{a_m b_i - a_i b_m}{x_{i,m}}}_{=\frac{1}{x_{i,m}}\cdot\left(a_m b_i - a_i b_m\right)}\right)$$

$$\cdot \underbrace{\frac{\prod\limits_{1\leq j<i\leq m-1} \left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)}{\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}}}_{=\frac{1}{\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}}\cdot\left(\prod\limits_{1\leq j<i\leq m-1}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)\right)}$$

$$= \left(\frac{1}{x_{m,m}}\right)^{m-1} \cdot \underbrace{\left(\prod_{j=1}^{m-1}\left(\frac{1}{x_{m,j}}\cdot\left(c_j d_m - c_m d_j\right)\right)\right)}_{=\left(\prod\limits_{j=1}^{m-1}\frac{1}{x_{m,j}}\right)\cdot\left(\prod\limits_{j=1}^{m-1}\left(c_j d_m - c_m d_j\right)\right)} \cdot \underbrace{\left(\prod_{i=1}^{m-1}\left(\frac{1}{x_{i,m}}\cdot\left(a_m b_i - a_i b_m\right)\right)\right)}_{=\left(\prod\limits_{i=1}^{m-1}\frac{1}{x_{i,m}}\right)\cdot\left(\prod\limits_{i=1}^{m-1}\left(a_m b_i - a_i b_m\right)\right)}$$

$$\cdot \underbrace{\frac{1}{\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}}}_{=\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}} \cdot \underbrace{\left(\prod_{1\leq j<i\leq m-1}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)\right)}_{=\left(\prod\limits_{1\leq j<i\leq m-1}\left(a_i b_j - a_j b_i\right)\right)\left(\prod\limits_{1\leq j<i\leq m-1}\left(c_j d_i - c_i d_j\right)\right)}$$

$$= \left(\frac{1}{x_{m,m}}\right)^{m-1} \cdot \left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right) \cdot \left(\prod_{j=1}^{m-1}\left(c_j d_m - c_m d_j\right)\right) \cdot \left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right) \cdot \left(\prod_{i=1}^{m-1}\left(a_m b_i - a_i b_m\right)\right)$$

$$\cdot \left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right) \cdot \left(\prod_{1\leq j<i\leq m-1}\left(a_i b_j - a_j b_i\right)\right)\left(\prod_{1\leq j<i\leq m-1}\left(c_j d_i - c_i d_j\right)\right)$$

$$= \left(\frac{1}{x_{m,m}}\right)^{m-1} \cdot \left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right) \cdot \left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right) \cdot \left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right)$$

$$\cdot \left(\prod_{i=1}^{m-1}\left(a_m b_i - a_i b_m\right)\right) \cdot \left(\prod_{1\leq j<i\leq m-1}\left(a_i b_j - a_j b_i\right)\right)$$

$$\cdot \left(\prod_{j=1}^{m-1}\left(c_j d_m - c_m d_j\right)\right) \cdot \left(\prod_{1\leq j<i\leq m-1}\left(c_j d_i - c_i d_j\right)\right). \tag{1160}$$

But

$$\begin{aligned}
&\underbrace{\prod_{(i,j)\in\{1,2,\ldots,m\}^2}}_{=\prod_{i\in\{1,2,\ldots,m\}}\prod_{j\in\{1,2,\ldots,m\}}} x_{i,j} \\
&= \underbrace{\prod_{i\in\{1,2,\ldots,m\}}}_{=\prod_{i=1}^{m}} \underbrace{\prod_{j\in\{1,2,\ldots,m\}}}_{=\prod_{j=1}^{m}} x_{i,j} = \prod_{i=1}^{m} \underbrace{\prod_{j=1}^{m} x_{i,j}}_{\substack{=\left(\prod_{j=1}^{m-1} x_{i,j}\right)\cdot x_{i,m} \\ \text{(here, we have split off the} \\ \text{factor for } j=m \text{ from the product)}}} \\
&= \prod_{i=1}^{m}\left(\left(\prod_{j=1}^{m-1} x_{i,j}\right)\cdot x_{i,m}\right) \\
&= \underbrace{\left(\prod_{i=1}^{m}\prod_{j=1}^{m-1} x_{i,j}\right)}_{\substack{=\left(\prod_{i=1}^{m-1}\prod_{j=1}^{m-1} x_{i,j}\right)\cdot\left(\prod_{j=1}^{m-1} x_{m,j}\right) \\ \text{(here, we have split off the factor for } i=m \text{ from the outer product)}}} \cdot \underbrace{\left(\prod_{i=1}^{m} x_{i,m}\right)}_{\substack{=\left(\prod_{i=1}^{m-1} x_{i,m}\right)\cdot x_{m,m} \\ \text{(here, we have split off the factor for } i=m \\ \text{from the product)}}} \\
&= \left(\underbrace{\prod_{i=1}^{m-1}}_{=\prod_{i\in\{1,2,\ldots,m-1\}}} \underbrace{\prod_{j=1}^{m-1}}_{=\prod_{j\in\{1,2,\ldots,m-1\}}} x_{i,j}\right)\cdot\left(\prod_{j=1}^{m-1} x_{m,j}\right)\cdot\left(\prod_{i=1}^{m-1} x_{i,m}\right)\cdot x_{m,m} \\
&= \left(\underbrace{\prod_{i\in\{1,2,\ldots,m-1\}}\prod_{j\in\{1,2,\ldots,m-1\}}}_{=\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}} x_{i,j}\right)\cdot\left(\prod_{j=1}^{m-1} x_{m,j}\right)\cdot\left(\prod_{i=1}^{m-1} x_{i,m}\right)\cdot x_{m,m} \\
&= \left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}\right)\cdot\left(\prod_{j=1}^{m-1} x_{m,j}\right)\cdot\left(\prod_{i=1}^{m-1} x_{i,m}\right)\cdot x_{m,m}.
\end{aligned}$$

Taking the (multiplicative) inverse of both sides of this equality, we obtain

$$\begin{aligned}
&\frac{1}{\prod\limits_{(i,j)\in\{1,2,\ldots,m\}^2} x_{i,j}} \\
&= \frac{1}{\left(\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}\right)\cdot\left(\prod\limits_{j=1}^{m-1} x_{m,j}\right)\cdot\left(\prod\limits_{i=1}^{m-1} x_{i,m}\right)\cdot x_{m,m}} \\
&= \underbrace{\frac{1}{\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2} x_{i,j}}}_{=\prod\limits_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}} \cdot \underbrace{\frac{1}{\prod\limits_{j=1}^{m-1} x_{m,j}}}_{=\prod\limits_{j=1}^{m-1}\frac{1}{x_{m,j}}} \cdot \underbrace{\frac{1}{\prod\limits_{i=1}^{m-1} x_{i,m}}}_{=\prod\limits_{i=1}^{m-1}\frac{1}{x_{i,m}}} \cdot \frac{1}{x_{m,m}} \\
&= \left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right)\cdot\left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right)\cdot\left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right)\cdot\frac{1}{x_{m,m}} \\
&= \frac{1}{x_{m,m}}\cdot\left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right)\cdot\left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right)\cdot\left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right).
\end{aligned}$$

Thus,

$$\frac{1}{x_{m,m}}\cdot\left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right)\cdot\left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right)\cdot\left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right) = \frac{1}{\prod\limits_{(i,j)\in\{1,2,\ldots,m\}^2} x_{i,j}}. \tag{1161}$$

Multiplying both sides of this equality by $x_{m,m}$, we obtain

$$\left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right)\cdot\left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right)\cdot\left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right) = x_{m,m}\cdot\frac{1}{\prod\limits_{(i,j)\in\{1,2,\ldots,m\}^2} x_{i,j}}. \tag{1162}$$

On the other hand,

$$\begin{aligned}
&\underbrace{\prod_{1\leq j<i\leq m}}_{=\prod_{i=1}^{m}\prod_{j=1}^{i-1}} \left(a_i b_j - a_j b_i\right) \\
&= \prod_{i=1}^{m}\prod_{j=1}^{i-1} \left(a_i b_j - a_j b_i\right) = \left(\underbrace{\prod_{i=1}^{m-1}\prod_{j=1}^{i-1}}_{=\prod_{1\leq j<i\leq m-1}} \left(a_i b_j - a_j b_i\right) \right) \cdot \prod_{j=1}^{m-1} \left(a_m b_j - a_j b_m\right) \\
&\qquad \text{(here, we have split off the factor for } i=m \text{ from the outer product)} \\
&= \left(\prod_{1\leq j<i\leq m-1} \left(a_i b_j - a_j b_i\right) \right) \cdot \underbrace{\left(\prod_{j=1}^{m-1} \left(a_m b_j - a_j b_m\right) \right)}_{\substack{=\prod_{i=1}^{m-1} (a_m b_i - a_i b_m) \\ \text{(here, we renamed the index } j \\ \text{as } i \text{ in the product)}}} \\
&= \left(\prod_{1\leq j<i\leq m-1} \left(a_i b_j - a_j b_i\right) \right) \cdot \left(\prod_{i=1}^{m-1} \left(a_m b_i - a_i b_m\right) \right) \\
&= \left(\prod_{i=1}^{m-1} \left(a_m b_i - a_i b_m\right) \right) \cdot \left(\prod_{1\leq j<i\leq m-1} \left(a_i b_j - a_j b_i\right) \right). \qquad (1163)
\end{aligned}$$

Also,

$$\begin{aligned}
&\underbrace{\prod_{1\leq j<i\leq m}}_{=\prod_{i=1}^{m}\prod_{j=1}^{i-1}} \left(c_j d_i - c_i d_j\right) \\
&= \prod_{i=1}^{m}\prod_{j=1}^{i-1} \left(c_j d_i - c_i d_j\right) = \left(\underbrace{\prod_{i=1}^{m-1}\prod_{j=1}^{i-1}}_{=\prod_{1\leq j<i\leq m-1}} \left(c_j d_i - c_i d_j\right) \right) \cdot \prod_{j=1}^{m-1} \left(c_j d_m - c_m d_j\right) \\
&\qquad \left(\text{here, we have split off the factor for } i=m \text{ from the outer product}\right) \\
&= \left(\prod_{1\leq j<i\leq m-1} \left(c_j d_i - c_i d_j\right) \right) \cdot \prod_{j=1}^{m-1} \left(c_j d_m - c_m d_j\right) \\
&= \left(\prod_{j=1}^{m-1} \left(c_j d_m - c_m d_j\right) \right) \cdot \left(\prod_{1\leq j<i\leq m-1} \left(c_j d_i - c_i d_j\right) \right). \qquad (1164)
\end{aligned}$$

Now, (1160) becomes

$$\det\left(\left(\frac{1}{x_{i,j}}\cdot\frac{1}{x_{m,m}}-\frac{1}{x_{i,m}}\cdot\frac{1}{x_{m,j}}\right)_{1\leq i\leq m-1,\ 1\leq j\leq m-1}\right)$$

$$= \underbrace{\left(\frac{1}{x_{m,m}}\right)^{m-1}}_{\substack{=\left(\frac{1}{x_{m,m}}\right)^{(m-2)+1} \\ =\left(\frac{1}{x_{m,m}}\right)^{m-2}\left(\frac{1}{x_{m,m}}\right)}} \cdot \underbrace{\left(\prod_{j=1}^{m-1}\frac{1}{x_{m,j}}\right)\cdot\left(\prod_{i=1}^{m-1}\frac{1}{x_{i,m}}\right)\cdot\left(\prod_{(i,j)\in\{1,2,\ldots,m-1\}^2}\frac{1}{x_{i,j}}\right)}_{\substack{=x_{m,m}\cdot\frac{1}{\prod_{(i,j)\in\{1,2,\ldots,m\}^2} x_{i,j}} \\ \text{(by (1162))}}}$$

$$\cdot \underbrace{\left(\prod_{i=1}^{m-1}\left(a_m b_i - a_i b_m\right)\right)\cdot\left(\prod_{1\leq j<i\leq m-1}\left(a_i b_j - a_j b_i\right)\right)}_{\substack{=\prod_{1\leq j<i\leq m}\left(a_i b_j - a_j b_i\right) \\ \text{(by (1163))}}}$$

$$\cdot \underbrace{\left(\prod_{j=1}^{m-1}\left(c_j d_m - c_m d_j\right)\right)\cdot\left(\prod_{1\leq j<i\leq m-1}\left(c_j d_i - c_i d_j\right)\right)}_{\substack{=\prod_{1\leq j<i\leq m}\left(c_j d_i - c_i d_j\right) \\ \text{(by (1164))}}}$$

$$= \underbrace{\left(\frac{1}{x_{m,m}}\right)^{m-2}}_{=\frac{1}{x_{m,m}^{m-2}}} \underbrace{\left(\frac{1}{x_{m,m}}\right)\cdot x_{m,m}}_{=1} \cdot \frac{1}{\prod_{(i,j)\in\{1,2,\ldots,m\}^2} x_{i,j}}$$

$$\cdot \underbrace{\left(\prod_{1\leq j<i\leq m}\left(a_i b_j - a_j b_i\right)\right)\cdot\left(\prod_{1\leq j<i\leq m}\left(c_j d_i - c_i d_j\right)\right)}_{=\prod_{1\leq j<i\leq m}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right)}$$

$$= \frac{1}{x_{m,m}^{m-2}}\cdot\frac{1}{\prod_{(i,j)\in\{1,2,\ldots,m\}^2} x_{i,j}}\cdot\prod_{1\leq j<i\leq m}\left(\left(a_i b_j - a_j b_i\right)\left(c_j d_i - c_i d_j\right)\right).$$

Multiplying both sides of this equality by $x_{m,m}^{m-2}$, we obtain

$$\begin{aligned}
& x_{m,m}^{m-2} \cdot \det\left(\left(\frac{1}{x_{i,j}} \cdot \frac{1}{x_{m,m}} - \frac{1}{x_{i,m}} \cdot \frac{1}{x_{m,j}} \right)_{1 \leq i \leq m-1,\ 1 \leq j \leq m-1} \right) \\
&= \underbrace{x_{m,m}^{m-2} \cdot \frac{1}{x_{m,m}^{m-2}}}_{=1} \cdot \frac{1}{\prod\limits_{(i,j) \in \{1,2,\ldots,m\}^2} x_{i,j}} \cdot \prod_{1 \leq j < i \leq m} \left(\left(a_i b_j - a_j b_i \right) \left(c_j d_i - c_i d_j \right) \right) \\
&= \frac{1}{\prod\limits_{(i,j) \in \{1,2,\ldots,m\}^2} x_{i,j}} \cdot \prod_{1 \leq j < i \leq m} \left(\left(a_i b_j - a_j b_i \right) \left(c_j d_i - c_i d_j \right) \right) \\
&= \frac{\prod\limits_{1 \leq j < i \leq m} \left(\left(a_i b_j - a_j b_i \right) \left(c_j d_i - c_i d_j \right) \right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,m\}^2} x_{i,j}}.
\end{aligned}$$

Compared with (1158), this yields

$$\det\left(\left(\frac{1}{x_{i,j}} \right)_{1 \leq i \leq m,\ 1 \leq j \leq m} \right) = \frac{\prod\limits_{1 \leq j < i \leq m} \left(\left(a_i b_j - a_j b_i \right) \left(c_j d_i - c_i d_j \right) \right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,m\}^2} x_{i,j}}.$$

In other words, (1156) holds for $k = m$. This completes the induction step. The induction proof of (1156) is thus complete.]

So we have proven (1156). We thus can apply (1156) to $k = n$, and obtain

$$\det\left(\left(\frac{1}{x_{i,j}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) = \frac{\prod\limits_{1 \leq j < i \leq n} \left(\left(a_i b_j - a_j b_i \right) \left(c_j d_i - c_i d_j \right) \right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,n\}^2} x_{i,j}}.$$

In light of (1155), this rewrites as

$$\det\left(\left(\frac{1}{a_i d_j - b_i c_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) = \frac{\prod\limits_{1 \leq j < i \leq n} \left(\left(a_i b_j - a_j b_i \right) \left(c_j d_i - c_i d_j \right) \right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,n\}^2} \left(a_i d_j - b_i c_j \right)}.$$

This proves Proposition 7.166. □

Now, let us solve Exercise 6.18.

Solution to Exercise 6.18. We have assumed that $x_i + y_j$ is an invertible element of $\mathbb{K}$ for every $(i,j) \in \{1,2,\ldots,n\}^2$. In other words, $x_i + y_j$ is an invertible element of $\mathbb{K}$ for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$. In other words, $x_i \cdot 1 - (-1) \cdot$

y_j is an invertible element of $\mathbb{K}$ for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ (because $\underbrace{x_i \cdot 1}_{=x_i} - \underbrace{(-1) \cdot y_j}_{=-y_j} = x_i - (-y_j) = x_i + y_j$ for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$). Hence, Proposition 7.166 (applied to x_i, -1, y_i and 1 instead of a_i, b_i, c_i and d_i) yields

$$\begin{aligned}
&\det\left(\left(\frac{1}{x_i \cdot 1 - (-1) \cdot y_j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) \\
&= \frac{\prod\limits_{1 \leq j < i \leq n} \left(\left(x_i \cdot (-1) - x_j \cdot (-1)\right)\left(y_j \cdot 1 - y_i \cdot 1\right)\right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,n\}^2} \left(x_i \cdot 1 - (-1) \cdot y_j\right)} \\
&= \left(\prod_{(i,j) \in \{1,2,\ldots,n\}^2} \underbrace{\left(x_i \cdot 1 - (-1) \cdot y_j\right)}_{=x_i+y_j}\right)^{-1} \\
&\qquad \cdot \prod_{1 \leq j < i \leq n} \left(\underbrace{\left(x_i \cdot (-1) - x_j \cdot (-1)\right)}_{=x_j - x_i} \underbrace{\left(y_j \cdot 1 - y_i \cdot 1\right)}_{=y_j - y_i}\right) \\
&= \left(\prod_{(i,j) \in \{1,2,\ldots,n\}^2} \left(x_i + y_j\right)\right)^{-1} \cdot \underbrace{\prod_{1 \leq j < i \leq n} \left(\left(x_j - x_i\right)\left(y_j - y_i\right)\right)}_{\substack{= \prod\limits_{1 \leq i < j \leq n} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right) \\ \text{(here, we have renamed the index } (j,i) \\ \text{as } (i,j) \text{ in the product)}}} \\
&= \left(\prod_{(i,j) \in \{1,2,\ldots,n\}^2} \left(x_i + y_j\right)\right)^{-1} \cdot \prod_{1 \leq i < j \leq n} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right) \\
&= \frac{\prod\limits_{1 \leq i < j \leq n} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,n\}^2} \left(x_i + y_j\right)}.
\end{aligned}$$

This rewrites as

$$\det\left(\left(\frac{1}{x_i + y_j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = \frac{\prod\limits_{1 \leq i < j \leq n} \left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right)}{\prod\limits_{(i,j) \in \{1,2,\ldots,n\}^2} \left(x_i + y_j\right)}$$

(because every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ satisfy $x_i \cdot 1 - (-1) \cdot y_j = x_i + y_j$). This solves Exercise 6.18. □

7.87. Solution to Exercise 6.20

Solution to Exercise 6.20. For every $n \times n$-matrix $\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, we have

$$\det\left(\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^{n} c_{\sigma(i),i} \qquad (1165)$$

[515]. Applied to $\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, this yields

$$\det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^{n} a_{\sigma(i),i}. \qquad (1166)$$

Now, let $k \in \{1,2,\ldots,n\}$. For every $\sigma \in S_n$, we have

$$\underbrace{\prod_{i=1}^{n}}_{=\prod_{i\in\{1,2,\ldots,n\}}} b_{\sigma(i)}^{\delta_{i,k}} = \prod_{i\in\{1,2,\ldots,n\}} b_{\sigma(i)}^{\delta_{i,k}} = \underbrace{b_{\sigma(k)}^{\delta_{k,k}}}_{\substack{=b_{\sigma(k)}^{1} \\ (\text{since } \delta_{k,k}=1)}} \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} \underbrace{b_{\sigma(i)}^{\delta_{i,k}}}_{\substack{=b_{\sigma(i)}^{0} \\ (\text{since } \delta_{i,k}=0 \\ (\text{since } i\neq k))}}$$

(here, we have split off the factor for $i = k$ from the product)

$$= \underbrace{b_{\sigma(k)}^{1}}_{=b_{\sigma(k)}} \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} \underbrace{b_{\sigma(i)}^{0}}_{=1} = b_{\sigma(k)} \underbrace{\prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq k}} 1}_{=1} = b_{\sigma(k)}. \qquad (1167)$$

---

[515] *Proof of (1165):* Let $\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be any $n \times n$-matrix. Then, Exercise 6.4 (applied to $A = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$) yields

$$\det\left(\left(\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)^T\right) = \det\left(\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

Hence,

$$\det\left(\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \det\left(\underbrace{\left(\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)^T}_{\substack{=\left(c_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \\ (\text{by the definition of the} \\ \text{transpose of a matrix})}}\right) = \det\left(\left(c_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$
$$= \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^{n} c_{\sigma(i),i}$$
$$\left(\begin{array}{c}\text{by (341), applied to } \left(c_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \text{ and } c_{j,i} \\ \text{instead of } A \text{ and } a_{i,j}\end{array}\right).$$

This proves (1165).

But we can apply (1165) to $\left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = \left(a_{i,j}b_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, and thus obtain

$$\det\left(\left(a_{i,j}b_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \sum_{\sigma\in S_n}(-1)^\sigma \underbrace{\prod_{i=1}^n\left(a_{\sigma(i),i}b_{\sigma(i)}^{\delta_{i,k}}\right)}_{=\left(\prod_{i=1}^n a_{\sigma(i),i}\right)\left(\prod_{i=1}^n b_{\sigma(i)}^{\delta_{i,k}}\right)}$$
$$= \sum_{\sigma\in S_n}(-1)^\sigma\left(\prod_{i=1}^n a_{\sigma(i),i}\right)\underbrace{\left(\prod_{i=1}^n b_{\sigma(i)}^{\delta_{i,k}}\right)}_{\substack{=b_{\sigma(k)}\\ \text{(by (1167))}}}$$
$$= \sum_{\sigma\in S_n}(-1)^\sigma\left(\prod_{i=1}^n a_{\sigma(i),i}\right)b_{\sigma(k)}. \qquad (1168)$$

Now, let us forget that we fixed k. We thus have proven (1168) for every $k\in\{1,2,\ldots,n\}$. On the other hand, for every $\sigma\in S_n$, we have

$$\underbrace{\sum_{k=1}^n}_{=\sum_{k\in\{1,2,\ldots,n\}}} b_{\sigma(k)} = \sum_{k\in\{1,2,\ldots,n\}} b_{\sigma(k)} = \underbrace{\sum_{k\in\{1,2,\ldots,n\}}}_{=\sum_{k=1}^n} b_k$$
$$\left(\begin{array}{c}\text{here, we have substituted } k \text{ for } \sigma(k) \text{ in the sum,}\\ \text{since } \sigma \text{ is a bijection } \{1,2,\ldots,n\}\to\{1,2,\ldots,n\}\\ \text{(since } \sigma \text{ is a permutation of } \{1,2,\ldots,n\} \text{ (since } \sigma\in S_n))\end{array}\right)$$
$$= \sum_{k=1}^n b_k = b_1+b_2+\cdots+b_n. \qquad (1169)$$

Now,

$$\sum_{k=1}^{n} \underbrace{\det\left(\left(a_{i,j} b_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{\substack{=\sum\limits_{\sigma\in S_n}(-1)^\sigma\left(\prod\limits_{i=1}^{n}a_{\sigma(i),i}\right)b_{\sigma(k)} \\ \text{(by (1168))}}}$$

$$= \sum_{k=1}^{n}\sum_{\sigma\in S_n}(-1)^\sigma\left(\prod_{i=1}^{n}a_{\sigma(i),i}\right)b_{\sigma(k)} = \sum_{\sigma\in S_n}(-1)^\sigma\left(\prod_{i=1}^{n}a_{\sigma(i),i}\right)\underbrace{\sum_{k=1}^{n}b_{\sigma(k)}}_{\substack{=b_1+b_2+\cdots+b_n \\ \text{(by (1169))}}}$$

$$= \sum_{\sigma\in S_n}(-1)^\sigma\left(\prod_{i=1}^{n}a_{\sigma(i),i}\right)(b_1+b_2+\cdots+b_n)$$

$$= (b_1+b_2+\cdots+b_n)\underbrace{\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^{n}a_{\sigma(i),i}}_{\substack{=\det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ \text{(by (1166))}}}$$

$$= (b_1+b_2+\cdots+b_n)\det\left(\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

This solves Exercise 6.20. □

7.88. Second solution to Exercise 6.16

Exercise 6.20 can be used to give a new (and simpler) solution to Exercise 6.16, suggested by Karthik Karnik:

Second solution to Exercise 6.16. Exercise 6.20 (applied to $a_{i,j} = x_i^{n-j}$ and $b_i = x_i$) shows that

$$\sum_{k=1}^{n}\det\left(\left(x_i^{n-j}x_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$
$$= (x_1+x_2+\cdots+x_n)\underbrace{\det\left(\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{\substack{=\prod\limits_{1\leq i<j\leq n}\left(x_i-x_j\right) \\ \text{(by Theorem 6.46 \textbf{(a)})}}}$$
$$= (x_1+x_2+\cdots+x_n)\prod_{1\leq i<j\leq n}\left(x_i-x_j\right). \tag{1170}$$

The left hand side of this equality is a sum of n determinants. We shall now show that $n-1$ of these determinants (namely, the ones that appear as addends for $k>1$ in the sum) are 0.

Indeed, every $k \in \{2, 3, \ldots, n\}$ satisfies

$$\det\left(\left(x_i^{n-j} x_i^{\delta_{j,k}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = 0 \qquad (1171)$$

[516]. Furthermore, every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfies

$$x_i^{n-j} x_i^{\delta_{j,1}} = \begin{cases} x_i^{n-j}, & \text{if } j > 1; \\ x_i^n, & \text{if } j = 1 \end{cases} \qquad (1173)$$

---

[516] *Proof of (1171):* Let $k \in \{2, 3, \ldots, n\}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j} x_i^{\delta_{j,k}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Let $h \in \{1, 2, \ldots, n\}$. Since $A = \left(x_i^{n-j} x_i^{\delta_{j,k}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, we have

$$(\text{the } (h,k)\text{-th entry of } A) = x_h^{n-k} \underbrace{x_h^{\delta_{k,k}}}_{\substack{=x_h^1 \\ (\text{since } \delta_{k,k}=1)}} = x_h^{n-k} x_h^1 = x_h^{n-k+1}. \qquad (1172)$$

On the other hand, $k - 1 \in \{1, 2, \ldots, n\}$ (since $k \in \{2, 3, \ldots, n\}$), and thus the $(h, k-1)$-th entry of A exists. Since $A = \left(x_i^{n-j} x_i^{\delta_{j,k}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, this entry satisfies

$$(\text{the } (h,k-1)\text{-th entry of } A) = x_h^{n-(k-1)} \underbrace{x_h^{\delta_{k-1,k}}}_{\substack{=x_h^0 \\ (\text{since } \delta_{k-1,k}=0)}} = x_h^{n-(k-1)} x_h^0 = x_h^{n-(k-1)} = x_h^{n-k+1}.$$

Comparing this with (1172), we obtain

$$(\text{the } (h,k)\text{-th entry of } A) = (\text{the } (h,k-1)\text{-th entry of } A).$$

Now, let us forget that we fixed h. We thus have shown that $(\text{the } (h,k)\text{-th entry of } A) = (\text{the } (h,k-1)\text{-th entry of } A)$ for every $h \in \{1, 2, \ldots, n\}$. In other words, the k-th column of A equals the $(k-1)$-st column of A. Thus, the matrix A has two equal columns (since $k - 1 \neq k$). Therefore, Exercise 6.7 **(f)** shows that $\det A = 0$. Since $A = \left(x_i^{n-j} x_i^{\delta_{j,k}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, this rewrites as $\det\left(\left(x_i^{n-j} x_i^{\delta_{j,k}}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right) = 0$. Qed.

[517]. Now, (1170) shows that

$$\left(x_1+x_2+\cdots+x_n\right)\prod_{1\leq i<j\leq n}\left(x_i-x_j\right)$$
$$=\sum_{k=1}^{n}\det\left(\left(x_i^{n-j}x_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$
$$=\det\left(\left(\underbrace{x_i^{n-j}x_i^{\delta_{j,1}}}_{\substack{=\begin{cases}x_i^{n-j}, & \text{if } j>1;\\ x_i^n, & \text{if } j=1\end{cases}\\ \text{(by (1173))}}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)+\sum_{k=2}^{n}\underbrace{\det\left(\left(x_i^{n-j}x_i^{\delta_{j,k}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{\substack{=0\\ \text{(by (1171))}}}$$

(here, we have split off the addend for $k=1$ from the sum)

$$=\det\left(\left(\begin{cases}x_i^{n-j}, & \text{if } j>1;\\ x_i^n, & \text{if } j=1\end{cases}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)+\underbrace{\sum_{k=2}^{n}0}_{=0}$$
$$=\det\left(\left(\begin{cases}x_i^{n-j}, & \text{if } j>1;\\ x_i^n, & \text{if } j=1\end{cases}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

This solves Exercise 6.16 again. □

7.89. Solution to Exercise 6.21

First solution to Exercise 6.21. The following solution will rely on Exercise 6.8 and 6.3. Since this is not the first time (nor the second) that we are using these exercises,

---

[517]*Proof of (1173):* Let $(i,j)\in\{1,2,\ldots,n\}^2$. We need to prove the equality (1173). To do so, it clearly satisfies to prove the following two claims:

Claim 1: If $j>1$, then $x_i^{n-j}x_i^{\delta_{j,1}}=x_i^{n-j}$.

Claim 2: If $j=1$, then $x_i^{n-j}x_i^{\delta_{j,1}}=x_i^n$.

Proof of Claim 1: Assume that $j>1$. Thus, $j\neq 1$, so that $\delta_{j,1}=0$, so that $x_i^{\delta_{j,1}}=x_i^0=1$. Hence, $x_i^{n-j}\underbrace{x_i^{\delta_{j,1}}}_{=1}=x_i^{n-j}$, and thus Claim 1 is proven.

Proof of Claim 2: Assume that $j=1$. Thus, $\delta_{j,1}=1$, so that $x_i^{\delta_{j,1}}=x_i^1$ and thus $x_i^{n-j}\underbrace{x_i^{\delta_{j,1}}}_{=x_i^1}=x_i^{n-j}x_i^1=x_i^{(n-j)+1}=x_i^{(n-1)+1}$ (since $j=1$), so that $x_i^{n-j}x_i^{\delta_{j,1}}=x_i^{(n-1)+1}=x_i^n$. This proves Claim 2.

Hence, (1173) is proven (since we have proven Claims 1 and 2).

I shall be brief.

Let A be the $(n+1)\times(n+1)$-matrix $\begin{pmatrix} x & a_1 & a_2 & \cdots & a_{n-1} & a_n \\ a_1 & x & a_2 & \cdots & a_{n-1} & a_n \\ a_1 & a_2 & x & \cdots & a_{n-1} & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 & a_2 & a_3 & \cdots & x & a_n \\ a_1 & a_2 & a_3 & \cdots & a_n & x \end{pmatrix}$. We need to prove that $\det A = \left(x + \sum\limits_{i=1}^{n} a_i\right) \prod\limits_{i=1}^{n} (x - a_i)$.

We perform the following operations on the matrix A (in this order):

- We subtract the 2-nd row from the 1-st row.
- We subtract the 3-rd row from the 2-nd row.
- ...
- We subtract the $(n+1)$-th row from the n-th row.

As we know from Exercise 6.8 **(a)**, these operations do not change the determinant of the matrix. Thus, if we denote by B the result of these operations, then $\det B = \det A$. On the other hand, it is easy to see that

$$B = \begin{pmatrix} x-a_1 & a_1-x & 0 & \cdots & 0 & 0 \\ 0 & x-a_2 & a_2-x & \cdots & 0 & 0 \\ 0 & 0 & x-a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x-a_n & a_n-x \\ a_1 & a_2 & a_3 & \cdots & a_n & x \end{pmatrix}.$$

(Here, for every $i \in \{1,2,\ldots,n\}$, the i-th row of B has i-th entry $x - a_i$ and $(i+1)$-th entry $a_i - x$; all other entries in this row are 0. The $(n+1)$-th row of B is $(a_1, a_2, \ldots, a_n, x)$.)

Next, we apply the following operations to the matrix B (in this order):

- We add the 1-st column to the 2-nd column.
- We add the 2-nd column to the 3-rd column.
- ...
- We add the n-th column to the $(n+1)$-th column.

(Notice that the order in which we are performing these operations forces their effects to accumulate; namely, at every step, the column that we are adding has already been modified by the previous step.) As we know from Exercise 6.8 **(b)**, these operations do not change the determinant of the matrix. Thus, if we denote by C the result of these operations, then $\det C = \det B = \det A$. On the other hand, it is easy to see that

$$C = \begin{pmatrix} x-a_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & x-a_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & x-a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x-a_n & 0 \\ a_1 & a_1+a_2 & a_1+a_2+a_3 & \cdots & a_1+a_2+\cdots+a_n & x+a_1+a_2+\cdots+a_n \end{pmatrix}.$$

This is a lower-triangular matrix. Thus, Exercise 6.3 shows that its determinant is the product of its diagonal entries. In other words,

$$\det C = (x-a_1)(x-a_2)\cdots(x-a_n)(x+a_1+a_2+\cdots+a_n) = \left(x+\sum_{i=1}^n a_i\right)\prod_{i=1}^n (x-a_i).$$

Compared with $\det C = \det A$, this yields $\det A = \left(x+\sum_{i=1}^n a_i\right)\prod_{i=1}^n (x-a_i)$. This solves Exercise 6.21. □

Second solution to Exercise 6.21. For any $(i,j) \in \{1,2,\ldots,n+1\}^2$, define an element $u_{i,j} \in \mathbb{K}$ by

$$u_{i,j} = \begin{cases} a_j, & \text{if } j<i; \\ x, & \text{if } j=i; \\ a_{j-1}, & \text{if } j>i \end{cases}.$$

This $u_{i,j}$ is well-defined[518]. Now, we define an $(n+1)\times(n+1)$-matrix U by $U = (u_{i,j})_{1\le i\le n+1,\ 1\le j\le n+1}$. Thus,

$$U = (u_{i,j})_{1\le i\le n+1,\ 1\le j\le n+1} = \begin{pmatrix} x & a_1 & a_2 & \cdots & a_{n-1} & a_n \\ a_1 & x & a_2 & \cdots & a_{n-1} & a_n \\ a_1 & a_2 & x & \cdots & a_{n-1} & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 & a_2 & a_3 & \cdots & x & a_n \\ a_1 & a_2 & a_3 & \cdots & a_n & x \end{pmatrix}.$$

[518] *Proof.* It is sufficient to check that a_j is well-defined when $j<i$, and that a_{j-1} is well-defined when $j>i$ (because x is always well-defined). But this is easy:

- If $j<i$, then $j\in\{1,2,\ldots,n\}$ (since $j<i\le n+1$ yields $j\le n$), and thus a_j is well-defined.
- If $j>i$, then $j\in\{2,3,\ldots,n+1\}$ (since $j>i\ge 1$ yields $j\ge 2$), and thus a_{j-1} is well-defined.

Our goal is now to prove that $\det U = \left(x + \sum_{i=1}^{n} a_i\right) \prod_{i=1}^{n} (x - a_i)$.

For any $(i,j) \in \{1,2,\ldots,n+1\}^2$, define an element $s_{i,j} \in \mathbb{K}$ by

$$s_{i,j} = \begin{cases} 1, & \text{if } i \leq j; \\ 0, & \text{if } i > j \end{cases}.$$

Now, we define an $(n+1) \times (n+1)$-matrix S by $S = \left(s_{i,j}\right)_{1\leq i\leq n+1,\ 1\leq j\leq n+1}$. The matrix S is upper-triangular[519]. Since the determinant of an upper-triangular matrix is the product of its diagonal entries, we thus obtain $\det S = 1 \cdot 1 \cdot \cdots \cdot 1 = 1$.

We extend the n-tuple $(a_1, a_2, \ldots, a_n) \in \mathbb{K}^n$ to an $(n+1)$-tuple $(a_1, a_2, \ldots, a_{n+1}) \in \mathbb{K}^{n+1}$ by setting $a_{n+1} = 0$. Thus, an element $a_k \in \mathbb{K}$ is defined for every $k \in \{1,2,\ldots,n+1\}$.

Now, for every $(i,j) \in \{1,2,\ldots,n+1\}^2$, we have

$$\sum_{k=1}^{n+1} u_{i,k} s_{k,j} = \sum_{k=1}^{j} a_k + s_{i,j}\left(x - a_j\right) \tag{1174}$$

[520].

For any two objects i and j, we define an element $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

For any $(i,j) \in \{1,2,\ldots,n+1\}^2$, define an element $c_{i,j} \in \mathbb{K}$ by

$$c_{i,j} = \delta_{i,j}\left(x - a_j\right) + \delta_{i,n+1} \sum_{k=1}^{j} a_k.$$

---

[519] It looks as follows: $S = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.

[520] *Proof of (1174):* Let $(i,j) \in \{1,2,\ldots,n+1\}^2$. We need to prove the equality (1174).

We have $(i,j) \in \{1,2,\ldots,n+1\}^2$. Thus, $1 \leq i \leq n+1$ and $1 \leq j \leq n+1$. Now,

$$\begin{aligned} \sum_{k=1}^{n+1} u_{i,k} s_{k,j} &= \sum_{k=1}^{j} u_{i,k} \underbrace{s_{k,j}}_{\substack{=1 \\ \text{(by the definition of} \\ s_{k,j}\text{, since } k\leq j)}} + \sum_{k=j+1}^{n+1} u_{i,k} \underbrace{s_{k,j}}_{\substack{=0 \\ \text{(by the definition of} \\ s_{k,j}\text{, since } k>j)}} \qquad (\text{since } 1 \leq j \leq n+1) \\ &= \sum_{k=1}^{j} u_{i,k} 1 + \underbrace{\sum_{k=j+1}^{n+1} u_{i,k} 0}_{=0} = \sum_{k=1}^{j} u_{i,k} 1 = \sum_{k=1}^{j} u_{i,k}. \end{aligned} \tag{1175}$$

Now, we must be in one of the following two cases:

Case 1: We have $i \leq j$.

Case 2: We have $i > j$.

Let us first consider Case 1. In this case, we have $i \leq j$. The definition of $s_{i,j}$ shows that $s_{i,j} = 1$

Let C be the $(n+1)\times(n+1)$-matrix defined by

$$C=\left(c_{i,j}\right)_{1\leq i\leq n+1,\ 1\leq j\leq n+1}.$$

(since $i\leq j$). Therefore,

$$\underbrace{\sum_{k=1}^{j}a_k}_{\substack{=\sum_{k=1}^{j-1}a_k+a_j \\ \text{(here, we have split off the} \\ \text{addend for } k=j \text{ from the sum)}}}+\underbrace{s_{i,j}}_{=1}\left(x-a_j\right)=\sum_{k=1}^{j-1}a_k+a_j+\left(x-a_j\right)=\sum_{k=1}^{j-1}a_k+x. \tag{1176}$$

We have $1\leq i\leq j$. Thus, $0\leq i-1\leq j-1$. Now, (1175) becomes

$$\begin{aligned}\sum_{k=1}^{n+1}u_{i,k}s_{k,j}=\sum_{k=1}^{j}u_{i,k}&=\underbrace{\sum_{k=1}^{i}u_{i,k}}_{\substack{=\sum_{k=1}^{i-1}u_{i,k}+u_{i,i} \\ \text{(here, we have split off the} \\ \text{addend for } k=i \text{ from the sum)}}}+\sum_{k=i+1}^{j}u_{i,k}\qquad(\text{since } 1\leq i\leq j)\\
&=\sum_{k=1}^{i-1}\underbrace{u_{i,k}}_{\substack{=a_k \\ \text{(by the definition of } u_{i,k}, \\ \text{since } k<i)}}+\underbrace{u_{i,i}}_{\substack{=x \\ \text{(by the definition of } u_{i,i}, \\ \text{since } i=i)}}+\sum_{k=i+1}^{j}\underbrace{u_{i,k}}_{\substack{=a_{k-1} \\ \text{(by the definition of } u_{i,k}, \\ \text{since } k>i)}}\\
&=\sum_{k=1}^{i-1}a_k+x+\underbrace{\sum_{k=i+1}^{j}a_{k-1}}_{\substack{=\sum_{k=i}^{j-1}a_k \\ \text{(here, we have substituted } k \text{ for } k-1 \text{ in the sum)}}}=\sum_{k=1}^{i-1}a_k+x+\sum_{k=i}^{j-1}a_k\\
&=\underbrace{\sum_{k=1}^{i-1}a_k+\sum_{k=i}^{j-1}a_k}_{\substack{=\sum_{k=1}^{j-1}a_k \\ (\text{since } 0\leq i-1\leq j-1)}}+x=\sum_{k=1}^{j-1}a_k+x.\end{aligned}$$

Compared with (1176), this yields

$$\sum_{k=1}^{n+1}u_{i,k}s_{k,j}=\sum_{k=1}^{j}a_k+s_{i,j}\left(x-a_j\right).$$

Thus, (1174) is proven in Case 1.

Let us now consider Case 2. In this case, we have $i>j$. Hence, $j<i$. The definition of $s_{i,j}$ therefore shows that $s_{i,j}=0$. Hence,

$$\sum_{k=1}^{j}a_k+\underbrace{s_{i,j}}_{=0}\left(x-a_j\right)=\sum_{k=1}^{j}a_k+\underbrace{0\left(x-a_j\right)}_{=0}=\sum_{k=1}^{j}a_k. \tag{1177}$$

Now, for every $(i,j) \in \{1,2,\ldots,n+1\}^2$, we have

$$\sum_{k=1}^{n+1} s_{i,k}c_{k,j} = \sum_{k=1}^{j} a_k + s_{i,j}\left(x - a_j\right) \tag{1178}$$

[521]. Thus, for every $(i,j) \in \{1,2,\ldots,n+1\}^2$, we have

$$\sum_{k=1}^{n+1} s_{i,k}c_{k,j} = \sum_{k=1}^{j} a_k + s_{i,j}\left(x - a_j\right) = \sum_{k=1}^{n+1} u_{i,k}s_{k,j} \tag{1179}$$

(by (1174)).

For every $i \in \{1,2,\ldots,n\}$, we have

$$c_{i,i} = x - a_i \tag{1180}$$

But (1175) becomes

$$\sum_{k=1}^{n+1} u_{i,k}s_{k,j} = \sum_{k=1}^{j} \underbrace{u_{i,k}}_{\substack{=a_k \\ \text{(by the definition of } u_{i,k}, \\ \text{since } k\leq j<i)}} = \sum_{k=1}^{j} a_k.$$

Compared with (1177), this yields

$$\sum_{k=1}^{n+1} u_{i,k}s_{k,j} = \sum_{k=1}^{j} a_k + s_{i,j}\left(x - a_j\right).$$

Thus, (1174) is proven in Case 2.

We have thus proven (1174) in each of the two Cases 1 and 2. Hence, (1174) always holds.

[521] *Proof of (1178):* Let $(i,j) \in \{1,2,\ldots,n+1\}^2$. We need to prove the equality (1178).

We have $(i,j) \in \{1,2,\ldots,n+1\}^2$. Thus, $1 \leq i \leq n+1$ and $1 \leq j \leq n+1$.

[522]. Also,

$$c_{n+1,n+1} = x + \sum_{i=1}^{n} a_i \qquad (1181)$$

The definition of $s_{i,n+1}$ yields $s_{i,n+1} = 1$ (since $i \le n+1$). Now,

$$\begin{aligned}
&\sum_{k=1}^{n+1} s_{i,k} c_{k,j} \\
&= \underbrace{\sum_{r=1}^{n+1}}_{=\sum_{r\in\{1,2,\ldots,n+1\}}} s_{i,r} \underbrace{c_{r,j}}_{\substack{=\delta_{r,j}(x-a_j)+\delta_{r,n+1}\sum_{k=1}^{j} a_k \\ \text{(by the definition of } c_{r,j})}} \qquad \text{(here, we renamed the summation index } k \text{ as } r) \\
&= \sum_{r\in\{1,2,\ldots,n+1\}} s_{i,r} \left(\delta_{r,j}\left(x-a_j\right) + \delta_{r,n+1}\sum_{k=1}^{j} a_k\right) \\
&= \underbrace{\sum_{r\in\{1,2,\ldots,n+1\}} s_{i,r}\delta_{r,j}\left(x-a_j\right)}_{\substack{=s_{i,j}\delta_{j,j}(x-a_j)+\sum_{r\in\{1,2,\ldots,n+1\};\ r\neq j} s_{i,r}\delta_{r,j}(x-a_j) \\ \text{(here, we have split off the addend for } r=j \text{ from the sum)}}} + \underbrace{\sum_{r\in\{1,2,\ldots,n+1\}} s_{i,r}\delta_{r,n+1}\sum_{k=1}^{j} a_k}_{\substack{=s_{i,n+1}\delta_{n+1,n+1}\sum_{k=1}^{j} a_k+\sum_{r\in\{1,2,\ldots,n+1\};\ r\neq n+1} s_{i,r}\delta_{r,n+1}\sum_{k=1}^{j} a_k \\ \text{(here, we have split off the addend for } r=n+1 \text{ from the sum)}}} \\
&= s_{i,j} \underbrace{\delta_{j,j}}_{\substack{=1 \\ (\text{since } j=j)}} \left(x-a_j\right) + \sum_{\substack{r\in\{1,2,\ldots,n+1\};\\ r\neq j}} s_{i,r} \underbrace{\delta_{r,j}}_{\substack{=0 \\ (\text{since } r\neq j)}} \left(x-a_j\right) \\
&\quad + \underbrace{s_{i,n+1}}_{=1} \underbrace{\delta_{n+1,n+1}}_{\substack{=1 \\ (\text{since } n+1=n+1)}} \sum_{k=1}^{j} a_k + \sum_{\substack{r\in\{1,2,\ldots,n+1\};\\ r\neq n+1}} s_{i,r} \underbrace{\delta_{r,n+1}}_{\substack{=0 \\ (\text{since } r\neq n+1)}} \sum_{k=1}^{j} a_k \\
&= s_{i,j}\left(x-a_j\right) + \underbrace{\sum_{\substack{r\in\{1,2,\ldots,n+1\};\\ r\neq j}} s_{i,r} 0 \left(x-a_j\right)}_{=0} + \sum_{k=1}^{j} a_k + \underbrace{\sum_{\substack{r\in\{1,2,\ldots,n+1\};\\ r\neq n+1}} s_{i,r} 0 \sum_{k=1}^{j} a_k}_{=0} \\
&= s_{i,j}\left(x-a_j\right) + \sum_{k=1}^{j} a_k = \sum_{k=1}^{j} a_k + s_{i,j}\left(x-a_j\right).
\end{aligned}$$

Thus, (1178) is proven.

[522] *Proof of (1180):* Let $i \in \{1,2,\ldots,n\}$. Thus, $i \neq n+1$, so that $\delta_{i,n+1} = 0$. Now, the definition of $c_{i,i}$ yields

$$c_{i,i} = \underbrace{\delta_{i,i}}_{\substack{=1 \\ (\text{since } i=i)}} \left(x-a_i\right) + \underbrace{\delta_{i,n+1}}_{=0} \sum_{k=1}^{i} a_k = \left(x-a_i\right) + 0 \underbrace{\sum_{k=1}^{i} a_k}_{=0} = x - a_i.$$

This proves (1180).

[523].

But we have $c_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,n+1\}^2$ satisfying $i < j$ [524]. Hence, Exercise 6.3 (applied to $n+1$, C and $c_{i,j}$ instead of n, A and $a_{i,j}$) shows that

$$\begin{aligned}
\det C &= c_{1,1}c_{2,2}\cdots c_{n+1,n+1} = \prod_{i=1}^{n+1} c_{i,i} \\
&= \left(\prod_{i=1}^{n} \underbrace{c_{i,i}}_{\substack{=x-a_i \\ \text{(by (1180))}}}\right) \underbrace{c_{n+1,n+1}}_{\substack{=x+\sum_{i=1}^{n} a_i \\ \text{(by (1181))}}} \\
&= \left(\prod_{i=1}^{n}(x-a_i)\right)\left(x+\sum_{i=1}^{n} a_i\right) = \left(x+\sum_{i=1}^{n} a_i\right)\prod_{i=1}^{n}(x-a_i). \qquad (1182)
\end{aligned}$$

But recall that $S = (s_{i,j})_{1\leq i\leq n+1,\ 1\leq j\leq n+1}$ and $C = (c_{i,j})_{1\leq i\leq n+1,\ 1\leq j\leq n+1}$. Hence, the definition of the product of two matrices shows that

$$SC = \left(\underbrace{\sum_{k=1}^{n+1} s_{i,k}c_{k,j}}_{\substack{=\sum_{k=1}^{n+1} u_{i,k}s_{k,j} \\ \text{(by (1179))}}}\right)_{1\leq i\leq n+1,\ 1\leq j\leq n+1} = \left(\sum_{k=1}^{n+1} u_{i,k}s_{k,j}\right)_{1\leq i\leq n+1,\ 1\leq j\leq n+1}. \qquad (1183)$$

On the other hand, $U = (u_{i,j})_{1\leq i\leq n+1,\ 1\leq j\leq n+1}$ and $S = (s_{i,j})_{1\leq i\leq n+1,\ 1\leq j\leq n+1}$.

---

[523] *Proof of (1181):* The definition of $c_{n+1,n+1}$ yields

$$\begin{aligned}
c_{n+1,n+1} &= \underbrace{\delta_{n+1,n+1}}_{\substack{=1 \\ \text{(since } n+1=n+1)}}(x-a_{n+1}) + \underbrace{\delta_{n+1,n+1}}_{\substack{=1 \\ \text{(since } n+1=n+1)}}\sum_{k=1}^{n+1} a_k = (x-a_{n+1}) + \sum_{k=1}^{n+1} a_k \\
&= x + \underbrace{\left(\sum_{k=1}^{n+1} a_k - a_{n+1}\right)}_{=\sum_{k=1}^{n} a_k} = x + \sum_{k=1}^{n} a_k = x + \sum_{i=1}^{n} a_i
\end{aligned}$$

(here, we have renamed the summation index k as i). This proves (1181).

[524] *Proof.* Let $(i,j) \in \{1,2,\ldots,n+1\}^2$ be such that $i < j$. Thus, $1 \leq i < j \leq n+1$, so that $i < n+1$. Hence, $i \neq n+1$ and thus $\delta_{i,n+1} = 0$. Also, $i < j$, so that $i \neq j$ and thus $\delta_{i,j} = 0$. Now, the definition of $c_{i,j}$ yields $c_{i,j} = \underbrace{\delta_{i,j}}_{=0}(x-a_j) + \underbrace{\delta_{i,n+1}}_{=0}\sum_{k=1}^{j} a_k = 0(x-a_j) + 0\sum_{k=1}^{j} a_k = 0$, qed.

Hence, the definition of the product of two matrices shows that

$$US = \left(\sum_{k=1}^{n+1} u_{i,k} s_{k,j} \right)_{1 \leq i \leq n+1,\ 1 \leq j \leq n+1} = SC \qquad \text{(by (1183))} .$$

Thus,

$$\det \left(\underbrace{US}_{=SC} \right) = \det (SC) = \underbrace{\det S}_{=1} \cdot \det C$$
$$\left(\begin{array}{c} \text{by Theorem 6.23, applied to } n+1 \text{, } S \text{ and } C \\ \text{instead of } n \text{, } A \text{ and } B \end{array} \right)$$
$$= \det C = \left(x + \sum_{i=1}^{n} a_i \right) \prod_{i=1}^{n} (x - a_i) \qquad \text{(by (1182))} .$$

Compared with

$$\det (US) = \det U \cdot \underbrace{\det S}_{=1}$$
$$\text{(by Theorem 6.23, applied to } n+1 \text{, } U \text{ and } S \text{ instead of } n \text{, } A \text{ and } B \text{)}$$
$$= \det U,$$

this yields

$$\det U = \left(x + \sum_{i=1}^{n} a_i \right) \prod_{i=1}^{n} (x - a_i) .$$

This solves Exercise 6.21. □

7.90. Solution to Exercise 6.22

Solution to Exercise 6.22. Let z be the permutation $\operatorname{cyc}_{n,n-1,\ldots,1}$ (where we are using the notations of Definition 5.37). Then, every $i \in \{1, 2, \ldots, n\}$ satisfies

$$z(i) = \begin{cases} i-1, & \text{if } i > 1; \\ n, & \text{if } i = 1 \end{cases} . \tag{1184}$$

(This follows easily from the definition of z.) In particular, $z(1) = n \neq 1$ (since $n > 1$), and thus $z \neq \operatorname{id}$. Every $i \in \{1, 2, \ldots, n\}$ satisfies

$$z(i) = \begin{cases} i-1, & \text{if } i > 1; \\ n, & \text{if } i = 1 \end{cases}$$
$$\equiv \begin{cases} i-1, & \text{if } i > 1; \\ i-1, & \text{if } i = 1 \end{cases} \qquad \left(\begin{array}{c} \text{since } n \equiv 0 = \underbrace{1}_{=i} -1 = i-1 \bmod n \\ \text{in the case when } i = 1 \end{array} \right)$$
$$= i - 1 \bmod n. \tag{1185}$$

Notice also that $z = \operatorname{cyc}_{n,n-1,\ldots,1}$, so that $(-1)^z = (-1)^{\operatorname{cyc}_{n,n-1,\ldots,1}} = (-1)^{n-1}$ (by Exercise 5.17 **(d)**, applied to $k = n$ and $(i_1, i_2, \ldots, i_k) = (n, n-1, \ldots, 1)$).

We recall the following simple fact: If p and q are two elements of $\{1, 2, \ldots, n\}$ such that $p \equiv q \operatorname{mod} n$, then $p = q$. We shall use this fact several times (tacitly) in the following arguments.

Now, I claim that if $\sigma \in S_n$ satisfies $\sigma \notin \{\operatorname{id}, z\}$, then

$$\left(\begin{array}{c} \text{there exists an } i \in \{1, 2, \ldots, n\} \text{ satisfying} \\ i \neq \sigma(i) \text{ and } i \not\equiv \sigma(i) + 1 \operatorname{mod} n \end{array} \right). \tag{1186}$$

[*Proof of (1186):* Let $\sigma \in S_n$ be such that $\sigma \notin \{\operatorname{id}, z\}$. Thus, $\sigma \neq \operatorname{id}$ and $\sigma \neq z$.

We need to prove (1186). Indeed, let us assume the contrary (for the sake of contradiction). Thus,

$$\text{every } i \in \{1, 2, \ldots, n\} \text{ satisfies either } i = \sigma(i) \text{ or } i \equiv \sigma(i) + 1 \operatorname{mod} n. \tag{1187}$$

[525]

There exists a $J \in \{1, 2, \ldots, n\}$ such that $\sigma(J) \neq J$ (since $\sigma \neq \operatorname{id}$). Let j be the smallest such J. Thus, $\sigma(j) \neq j$, but every $J < j$ satisfies $\sigma(J) = J$.

Applying (1187) to $i = j$, we see that either $j = \sigma(j)$ or $j \equiv \sigma(j) + 1 \operatorname{mod} n$. Since $j = \sigma(j)$ cannot hold (because we have $\sigma(j) \neq j$), we thus have $j \equiv \sigma(j) + 1 \operatorname{mod} n$. In other words, $\sigma(j) \equiv j - 1 \operatorname{mod} n$.

We have $j = 1$ [526]. Thus, $\sigma\left(\underbrace{1}_{=j}\right) = \sigma(j) \equiv \underbrace{j}_{=1} - 1 = 0 \equiv n \operatorname{mod} n$. Since both $\sigma(1)$ and n belong to $\{1, 2, \ldots, n\}$, this shows that $\sigma(1) = n$.

There exists a $K \in \{1, 2, \ldots, n\}$ such that $\sigma(K) = K$ [527]. Let k be the largest such K. Thus, $\sigma(k) = k$, but every $K > k$ satisfies $\sigma(K) \neq K$.

We have $n \neq 1$ and thus $\sigma(n) \neq \sigma(1)$ (since σ is injective). Thus, $\sigma(n) \neq \sigma(1) = n$.

---

[525] I use the words "either"/"or" in a non-exclusive meaning (i.e., when I say "either $\mathcal{A}$ or $\mathcal{B}$", I mean to include also the case when both $\mathcal{A}$ and $\mathcal{B}$ hold simultaneously), but here it does not matter (because we cannot have $i = \sigma(i)$ and $i \equiv \sigma(i) + 1 \operatorname{mod} n$ at the same time).

[526] *Proof.* Assume the contrary. Thus, $j \neq 1$, so that $j > 1$. Hence, $j - 1 \in \{1, 2, \ldots, n\}$. Also, $j - 1 < j$. Hence, $\sigma(j-1) = j - 1$ (since every $J < j$ satisfies $\sigma(J) = J$). Thus, $\sigma(j-1) = j - 1 \equiv \sigma(j) \operatorname{mod} n$. Since both $\sigma(j-1)$ and $\sigma(j)$ are elements of $\{1, 2, \ldots, n\}$, this shows that $\sigma(j-1) = \sigma(j)$, and thus $j - 1 = j$ (since σ is injective). But this is absurd. This contradiction shows that our assumption was wrong, qed.

[527] *Proof.* We have $\sigma \neq z$. Hence, there exists an $i \in \{1, 2, \ldots, n\}$ such that $\sigma(i) \neq z(i)$. Consider this i.

If we had $i \equiv \sigma(i) + 1 \operatorname{mod} n$, then we would have $\sigma(i) \equiv i - 1 \equiv z(i) \operatorname{mod} n$ (by (1185)), which would entail $\sigma(i) = z(i)$ (since both $\sigma(i)$ and $z(i)$ belong to $\{1, 2, \ldots, n\}$); but this would contradict $\sigma(i) \neq z(i)$. Hence, we cannot have $i \equiv \sigma(i) + 1 \operatorname{mod} n$.

We have either $i = \sigma(i)$ or $i \equiv \sigma(i) + 1 \operatorname{mod} n$ (because of (1187)). Thus, $i = \sigma(i)$ (since we cannot have $i \equiv \sigma(i) + 1 \operatorname{mod} n$). Hence, there exists a $K \in \{1, 2, \ldots, n\}$ such that $\sigma(K) = K$ (namely, $K = i$). Qed.

We cannot have $k = n$ (because otherwise, we would have $\sigma\left(\underbrace{n}_{=k}\right) = \sigma(k) = k = n$, which would contradict $\sigma(n) \neq n$). Thus, $k < n$. Hence, $k+1 \in \{1, 2, \ldots, n\}$. Therefore, $\sigma(k+1) \neq k+1$ (since every $K > k$ satisfies $\sigma(K) \neq K$, and since $k+1 > k$). Now, applying (1187) to $i = k+1$, we conclude that either $k+1 = \sigma(k+1)$ or $k+1 \equiv \sigma(k+1) + 1 \mod n$. Since we cannot have $k+1 = \sigma(k+1)$ (because $\sigma(k+1) \neq k+1$), we thus must have $k+1 \equiv \sigma(k+1) + 1 \mod n$. In other words, $k \equiv \sigma(k+1) \mod n$. Hence, $k = \sigma(k+1)$ (since both k and $\sigma(k+1)$ belong to $\{1, 2, \ldots, n\}$), so that $\sigma(k+1) = k = \sigma(k)$. Since σ is injective, this yields $k+1 = k$, which is absurd. This contradiction shows that our assumption was wrong. Hence, (1186) is proven.]

Let us now write our matrix A in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Then,

$$(a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n} = A = \left(\begin{cases} a_j, & \text{if } i = j; \\ b_j, & \text{if } i \equiv j+1 \mod n; \\ 0, & \text{otherwise} \end{cases} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

In other words, we have

$$a_{i,j} = \begin{cases} a_j, & \text{if } i = j; \\ b_j, & \text{if } i \equiv j+1 \mod n; \\ 0, & \text{otherwise} \end{cases} \tag{1188}$$

for every $(i,j) \in \{1, 2, \ldots, n\}^2$.

For every $i \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned} a_{i,i} &= \begin{cases} a_i, & \text{if } i = i; \\ b_i, & \text{if } i \equiv i+1 \mod n; \\ 0, & \text{otherwise} \end{cases} \qquad \left(\text{by (1188), applied to } (i,i) \text{ instead of } (i,j)\right) \\ &= a_i \qquad (\text{since } i = i) \end{aligned} \tag{1189}$$

and

$$\begin{aligned} a_{i,z(i)} &= \begin{cases} a_{z(i)}, & \text{if } i = z(i); \\ b_{z(i)}, & \text{if } i \equiv z(i) + 1 \mod n; \\ 0, & \text{otherwise} \end{cases} \\ &\qquad \left(\text{by (1188), applied to } (i, z(i)) \text{ instead of } (i,j)\right) \\ &= b_{z(i)} \qquad (\text{since } i \equiv z(i) + 1 \mod n \text{ (by (1185))}). \end{aligned} \tag{1190}$$

It is now easy to see that if $\sigma \in S_n$ satisfies $\sigma \notin \{\mathrm{id}, z\}$, then

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = 0 \tag{1191}$$

[528].

Now, (341) becomes

$$\begin{aligned}
\det A &= \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} \\
&= \underbrace{(-1)^{\mathrm{id}}}_{=1} \prod_{i=1}^{n} \underbrace{a_{i,\mathrm{id}(i)}}_{=a_{i,i}} + \underbrace{(-1)^{z}}_{=(-1)^{n-1}} \prod_{i=1}^{n} a_{i,z(i)} + \sum_{\substack{\sigma \in S_n; \\ \sigma \notin \{\mathrm{id}, z\}}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{=0 \\ \text{(by (1191))}}} \\
&\qquad \left(\begin{array}{c} \text{here, we have split off the addends for } \sigma = \mathrm{id} \text{ and} \\ \text{for } \sigma = z \text{ from the sum (since } z \neq \mathrm{id}\text{)} \end{array} \right) \\
&= \prod_{i=1}^{n} a_{i,i} + (-1)^{n-1} \prod_{i=1}^{n} a_{i,z(i)} + \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma \notin \{\mathrm{id}, z\}}} (-1)^{\sigma} 0}_{=0} \\
&= \prod_{i=1}^{n} \underbrace{a_{i,i}}_{\substack{=a_i \\ \text{(by (1189))}}} + (-1)^{n-1} \prod_{i=1}^{n} \underbrace{a_{i,z(i)}}_{\substack{=b_{z(i)} \\ \text{(by (1190))}}} \\
&= \prod_{i=1}^{n} a_i + (-1)^{n-1} \underbrace{\prod_{i=1}^{n} b_{z(i)}}_{\substack{=\prod_{i=1}^{n} b_i \\ \text{(here, we have substituted } i \\ \text{for } z(i) \text{ in the product,} \\ \text{since } z:\{1,2,\ldots,n\} \to \{1,2,\ldots,n\} \\ \text{is a bijection)}}} = \underbrace{\prod_{i=1}^{n} a_i}_{=a_1 a_2 \cdots a_n} + (-1)^{n-1} \underbrace{\prod_{i=1}^{n} b_i}_{=b_1 b_2 \cdots b_n} \\
&= a_1 a_2 \cdots a_n + (-1)^{n-1} b_1 b_2 \cdots b_n .
\end{aligned}$$

This solves Exercise 6.22. □

---

[528] *Proof of (1191):* Let $\sigma \in S_n$ be such that $\sigma \notin \{\mathrm{id}, z\}$. Thus, there exists an $i \in \{1, 2, \ldots, n\}$ satisfying $i \neq \sigma(i)$ and $i \not\equiv \sigma(i) + 1 \mod n$ (because of (1186)). Consider this i.

From (1188) (applied to $(i, \sigma(i))$ instead of (i, j)), we obtain

$$a_{i,\sigma(i)} = \begin{cases} a_{\sigma(i)}, & \text{if } i = \sigma(i); \\ b_{\sigma(i)}, & \text{if } i \equiv \sigma(i) + 1 \mod n; \\ 0, & \text{otherwise} \end{cases} = 0$$

(since $i \neq \sigma(i)$ and $i \not\equiv \sigma(i) + 1 \mod n$).

Now, let us forget that we fixed i. We thus have shown that there exists an $i \in \{1, 2, \ldots, n\}$ satisfying $a_{i,\sigma(i)} = 0$. In other words, at least one factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ equals 0. Therefore, the whole product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ equals 0. This proves (1191).

7.91. Solution to Exercise 6.23

Before we come to the solution of Exercise 6.23, let us first state some simple lemmas:

Lemma 7.167. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in \mathbb{N}$ and $q \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} \in \mathbb{K}^{n\times m}$, $B = \left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq p} \in \mathbb{K}^{m\times p}$ and $C = \left(c_{i,j}\right)_{1\leq i\leq p,\ 1\leq j\leq q} \in \mathbb{K}^{p\times q}$. Then,

$$ABC = \left(\sum_{k=1}^{m}\sum_{\ell=1}^{p} a_{i,k}b_{k,\ell}c_{\ell,j}\right)_{1\leq i\leq n,\ 1\leq j\leq q}.$$

Proof of Lemma 7.167. We have $B = \left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq p}$ and $C = \left(c_{i,j}\right)_{1\leq i\leq p,\ 1\leq j\leq q}$. Thus, the definition of the product BC yields

$$BC = \left(\underbrace{\sum_{k=1}^{p} b_{i,k}c_{k,j}}_{\substack{=\sum_{\ell=1}^{p} b_{i,\ell}c_{\ell,j} \\ \text{(here, we renamed the} \\ \text{summation index } k \text{ as } \ell\text{)}}}\right)_{1\leq i\leq m,\ 1\leq j\leq q} = \left(\sum_{\ell=1}^{p} b_{i,\ell}c_{\ell,j}\right)_{1\leq i\leq m,\ 1\leq j\leq q}.$$

Now, we have $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ and $BC = \left(\sum_{\ell=1}^{p} b_{i,\ell}c_{\ell,j}\right)_{1\leq i\leq m,\ 1\leq j\leq q}$. Hence, the definition of the product $A(BC)$ yields

$$A(BC) = \left(\sum_{k=1}^{m}\underbrace{a_{i,k}\left(\sum_{\ell=1}^{p} b_{k,\ell}c_{\ell,j}\right)}_{=\sum_{\ell=1}^{p} a_{i,k}b_{k,\ell}c_{\ell,j}}\right)_{1\leq i\leq n,\ 1\leq j\leq q} = \left(\sum_{k=1}^{m}\sum_{\ell=1}^{p} a_{i,k}b_{k,\ell}c_{\ell,j}\right)_{1\leq i\leq n,\ 1\leq j\leq q}.$$

Thus,

$$ABC = A(BC) = \left(\sum_{k=1}^{m}\sum_{\ell=1}^{p} a_{i,k}b_{k,\ell}c_{\ell,j}\right)_{1\leq i\leq n,\ 1\leq j\leq q}.$$

This proves Lemma 7.167. □

Lemma 7.168. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. Let $S = \left(s_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$ be an $n \times m$-matrix. Then,

$$S^T A S = \left(\sum_{(k,\ell) \in \{1,2,\ldots,n\}^2} s_{k,i} s_{\ell,j} a_{k,\ell} \right)_{1 \leq i \leq m,\ 1 \leq j \leq m}.$$

Proof of Lemma 7.168. We have $S = \left(s_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Thus, $S^T = \left(s_{j,i}\right)_{1 \leq i \leq m,\ 1 \leq j \leq n}$ (by the definition of S^T). Recall also that $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ and $S = \left(s_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Hence, Lemma 7.167 (applied to m, n, n, m, S^T, A, S, $s_{j,i}$, $a_{i,j}$ and $s_{i,j}$ instead of n, m, p, q, A, B, C, $a_{i,j}$, $b_{i,j}$ and $c_{i,j}$) yields

$$\begin{aligned}
S^T A S &= \left(\underbrace{\sum_{k=1}^{n}}_{= \sum_{k \in \{1,2,\ldots,n\}}} \underbrace{\sum_{\ell=1}^{n}}_{= \sum_{\ell \in \{1,2,\ldots,n\}}} s_{k,i} \underbrace{a_{k,\ell} s_{\ell,j}}_{= s_{\ell,j} a_{k,\ell}} \right)_{1 \leq i \leq m,\ 1 \leq j \leq m} \\
&= \left(\underbrace{\sum_{k \in \{1,2,\ldots,n\}} \sum_{\ell \in \{1,2,\ldots,n\}}}_{= \sum_{(k,\ell) \in \{1,2,\ldots,n\}^2}} s_{k,i} s_{\ell,j} a_{k,\ell} \right)_{1 \leq i \leq m,\ 1 \leq j \leq m} \\
&= \left(\sum_{(k,\ell) \in \{1,2,\ldots,n\}^2} s_{k,i} s_{\ell,j} a_{k,\ell} \right)_{1 \leq i \leq m,\ 1 \leq j \leq m}.
\end{aligned}$$

This proves Lemma 7.168. □

Lemma 7.169. Let $n \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an alternating $n \times n$-matrix.

(a) Every $i \in \{1, 2, \ldots, n\}$ satisfies $a_{i,i} = 0$.

(b) Every $(i, j) \in \{1, 2, \ldots, n\}^2$ satisfy $a_{i,j} = -a_{j,i}$.

Proof of Lemma 7.169. Recall that the matrix A is alternating if and only if it satisfies $A^T = -A$ and ($a_{i,i} = 0$ for all $i \in \{1, 2, \ldots, n\}$) (by the definition of "alternating"). Hence, the matrix A satisfies $A^T = -A$ and ($a_{i,i} = 0$ for all $i \in \{1, 2, \ldots, n\}$) (since A is alternating).

We have $\left(a_{i,i}=0 \text{ for all } i\in\{1,2,\ldots,n\}\right)$. In other words, every $i\in\{1,2,\ldots,n\}$ satisfies $a_{i,i}=0$. This proves Lemma 7.169 **(a)**.

(b) We have $A=\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, and thus $A^T=\left(a_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of A^T). Hence,

$$\left(a_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=A^T=-\underbrace{A}_{=\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}}=-\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=\left(-a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

In other words,

$$a_{j,i}=-a_{i,j}\qquad\text{for every } (i,j)\in\{1,2,\ldots,n\}^2. \tag{1192}$$

Now, let $(i,j)\in\{1,2,\ldots,n\}^2$. Hence, (1192) yields $a_{j,i}=-a_{i,j}$. In other words, $a_{i,j}=-a_{j,i}$. This proves Lemma 7.169 **(b)**. □

Solution to Exercise 6.23. Let B be the $m\times m$-matrix S^TAS. Thus, $B=S^TAS$. Write the $m\times m$-matrix B in the form $B=\left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$.

Write the $n\times n$-matrix A in the form $A=\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Write the $n\times m$-matrix S in the form $S=\left(s_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$. Then,

$$B=S^TAS=\left(\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2}s_{k,i}s_{\ell,j}a_{k,\ell}\right)_{1\leq i\leq m,\ 1\leq j\leq m} \tag{1193}$$

(by Lemma 7.168).

But $B=\left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$, so that

$$\left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}=B=\left(\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2}s_{k,i}s_{\ell,j}a_{k,\ell}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$$

(by (1193)). In other words,

$$b_{i,j}=\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2}s_{k,i}s_{\ell,j}a_{k,\ell}\qquad\text{for every } (i,j)\in\{1,2,\ldots,m\}^2. \tag{1194}$$

But

$$\left(b_{j,i}\right)_{1\leq i\leq m,\ 1\leq j\leq m}=\left(-b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$$

[529]. But $B = \left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$, and thus $B^T = \left(b_{j,i}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$ (by the definition of B^T). Hence,

$$B^T = \left(b_{j,i}\right)_{1\leq i\leq m,\ 1\leq j\leq m} = \left(-b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}.$$

Comparing this with

$$-\underbrace{B}_{=\left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}} = -\left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m} = \left(-b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m},$$

we obtain $B^T = -B$.

[529] *Proof:* Every $(i,j) \in \{1,2,\ldots,m\}^2$ satisfies

$$\begin{aligned} b_{j,i} &= \sum_{(k,\ell)\in\{1,2,\ldots,n\}^2} s_{k,j}s_{\ell,i}a_{k,\ell} \qquad \text{(by (1194) (applied to } (j,i) \text{ instead of } (i,j)\text{))} \\ &= \underbrace{\sum_{(\ell,k)\in\{1,2,\ldots,n\}^2}}_{=\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2}} \underbrace{s_{\ell,j}s_{k,i}}_{=s_{k,i}s_{\ell,j}} \underbrace{a_{\ell,k}}_{\substack{=-a_{k,\ell} \\ \text{(by Lemma 7.169 \textbf{(b)}} \\ \text{(applied to } (\ell,k) \\ \text{instead of } (i,j)))}} \qquad \left(\begin{array}{c}\text{here, we have renamed the} \\ \text{summation index } (k,\ell) \text{ as } (\ell,k)\end{array}\right) \\ &= \sum_{(k,\ell)\in\{1,2,\ldots,n\}^2} s_{k,i}s_{\ell,j}\left(-a_{k,\ell}\right) = -\underbrace{\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2} s_{k,i}s_{\ell,j}a_{k,\ell}}_{\substack{=b_{i,j} \\ \text{(by (1194))}}} = -b_{i,j}. \end{aligned}$$

In other words, $\left(b_{j,i}\right)_{1\leq i\leq m,\ 1\leq j\leq m} = \left(-b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$.

Let $i \in \{1, 2, \ldots, m\}$. Then, (1194) (applied to (i, i) instead of (i, j)) yields

$$\begin{aligned}
b_{i,i} &= \sum_{(k,\ell) \in \{1,2,\ldots,n\}^2} s_{k,i} s_{\ell,i} a_{k,\ell} \\
&= \sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k<\ell}} s_{k,i} s_{\ell,i} a_{k,\ell} + \sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k=\ell}} s_{k,i} s_{\ell,i} \underbrace{a_{k,\ell}}_{\substack{= a_{\ell,\ell} \\ (\text{since } k=\ell)}} \\
&\qquad + \sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k>\ell}} \underbrace{s_{k,i} s_{\ell,i}}_{= s_{\ell,i} s_{k,i}} \underbrace{a_{k,\ell}}_{\substack{= -a_{\ell,k} \\ (\text{by Lemma 7.169 } \textbf{(b)} \\ (\text{applied to } (k,\ell) \text{ instead of } (i,j)))}} \\
&\qquad \left(\begin{array}{c} \text{since each } (k,\ell) \in \{1,2,\ldots,n\}^2 \text{ satisfies exactly one of} \\ \text{the three assertions } (k<\ell),\ (k=\ell) \text{ and } (k>\ell) \end{array} \right) \\
&= \underbrace{\sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k<\ell}} s_{k,i} s_{\ell,i} a_{k,\ell}}_{\substack{= \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ \ell<k}} s_{\ell,i} s_{k,i} a_{\ell,k} \\ (\text{here, we have renamed the} \\ \text{summation index } (k,\ell) \text{ as } (\ell,k))}} + \sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k=\ell}} s_{k,i} s_{\ell,i} \underbrace{a_{\ell,\ell}}_{\substack{=0 \\ (\text{by Lemma 7.169 } \textbf{(a)} \\ (\text{applied to } \ell \text{ instead of } i))}} \\
&\qquad + \underbrace{\sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k>\ell}} s_{\ell,i} s_{k,i} \left(-a_{\ell,k}\right)}_{= -\sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k>\ell}} s_{\ell,i} s_{k,i} a_{\ell,k}} \\
&= \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ \ell<k}} s_{\ell,i} s_{k,i} a_{\ell,k} + \underbrace{\sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k=\ell}} s_{k,i} s_{\ell,i} 0}_{=0} + \left(- \sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k>\ell}} s_{\ell,i} s_{k,i} a_{\ell,k} \right) \\
&= \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ \ell<k}} s_{\ell,i} s_{k,i} a_{\ell,k} - \underbrace{\sum_{\substack{(k,\ell) \in \{1,2,\ldots,n\}^2; \\ k>\ell}}}_{\substack{= \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ k>\ell}} = \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ \ell<k}} \\ (\text{because for every } (\ell,k) \in \{1,2,\ldots,n\}^2, \\ \text{the statement } (k>\ell) \text{ is equivalent to } (\ell<k))}} s_{\ell,i} s_{k,i} a_{\ell,k} \\
&= \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ \ell<k}} s_{\ell,i} s_{k,i} a_{\ell,k} - \sum_{\substack{(\ell,k) \in \{1,2,\ldots,n\}^2; \\ \ell<k}} s_{\ell,i} s_{k,i} a_{\ell,k} = 0.
\end{aligned}$$

Now, forget that we fixed i. We thus have shown that ($b_{i,i} = 0$ for all $i \in \{1, 2, \ldots, m\}$).

Now, recall that $B = \left(b_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq m}$. Hence, the $m \times m$-matrix B is alternating if and only if it satisfies $B^T = -B$ and ($b_{i,i} = 0$ for all $i \in \{1, 2, \ldots, m\}$) (by the definition of "alternating"). Thus, the $m \times m$-matrix B is alternating (since it satisfies $B^T = -B$ and ($b_{i,i} = 0$ for all $i \in \{1, 2, \ldots, m\}$)). Since $B = S^T AS$, this rewrites as follows: The $m \times m$-matrix $S^T AS$ is alternating. This solves Exercise 6.23. □

7.92. Solution to Exercise 6.24

Before we solve Exercise 6.24, let us show a combinatorial lemma:

Lemma 7.170. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ be such that $\sigma^{-1} = \sigma$. Let $X = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) = i\}$. Then,

$$|X| \equiv n \bmod 2.$$

Proof of Lemma 7.170. Define two further sets Y and Z by

$$Y = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) < i\}$$

and

$$Z = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) > i\}.$$

Every $j \in Y$ satisfies $\sigma(j) \in Z$ [530]. Thus, we can define a map $\alpha : Y \to Z$ by

$$(\alpha(j) = \sigma(j) \qquad \text{for every } j \in Y).$$

Consider this map α. Similarly, define a map $\beta : Z \to Y$ by

$$(\beta(j) = \sigma(j) \qquad \text{for every } j \in Z).$$

[530] *Proof.* Let $j \in Y$. Then,

$$j \in Y = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) < i\}.$$

In other words, j is an element i of $\{1, 2, \ldots, n\}$ satisfying $\sigma(i) < i$. In other words, j is an element of $\{1, 2, \ldots, n\}$ and satisfies $\sigma(j) < j$.

Clearly, $\sigma(j)$ is an element of $\{1, 2, \ldots, n\}$ and satisfies $\sigma(\sigma(j)) > \sigma(j)$ (since $\sigma(j) < j = \underbrace{\sigma^{-1}}_{=\sigma}(\sigma(j)) = \sigma(\sigma(j))$). In other words, $\sigma(j)$ is an element i of $\{1, 2, \ldots, n\}$ satisfying $\sigma(i) > i$. In other words, $\sigma(j) \in \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) > i\}$. This rewrites as $\sigma(j) \in Z$ (since $Z = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) > i\}$). Qed.

We have $\alpha \circ \beta = \operatorname{id}$ [531] and $\beta \circ \alpha = \operatorname{id}$ [532]. Thus, the two maps α and β are mutually inverse. Hence, the map α is invertible, i.e., is a bijection. Thus, there exists a bijection $Y \to Z$ (namely, the map α). Hence, $|Y| = |Z|$.

Now,

$$\sum_{i \in \{1,2,\ldots,n\}} 1 = \underbrace{|\{1,2,\ldots,n\}|}_{=n} \cdot 1 = n \cdot 1 = n.$$

Hence,

$$\begin{aligned}
n &= \sum_{i \in \{1,2,\ldots,n\}} 1 \\
&= \underbrace{\sum_{\substack{i \in \{1,2,\ldots,n\};\\ \sigma(i)<i}}}_{\substack{=\sum_{i\in Y}\\ (\text{since } \{i \in \{1,2,\ldots,n\} \mid \sigma(i)<i\}=Y)}} 1 + \underbrace{\sum_{\substack{i \in \{1,2,\ldots,n\};\\ \sigma(i)=i}}}_{\substack{=\sum_{i\in X}\\ (\text{since } \{i \in \{1,2,\ldots,n\} \mid \sigma(i)=i\}=X)}} 1 \\
&\qquad + \underbrace{\sum_{\substack{i \in \{1,2,\ldots,n\};\\ \sigma(i)>i}}}_{\substack{=\sum_{i\in Z}\\ (\text{since } \{i \in \{1,2,\ldots,n\} \mid \sigma(i)>i\}=Z)}} 1 \\
&\qquad \left(\begin{array}{c}\text{since each } i \in \{1,2,\ldots,n\} \text{ satisfies exactly one of the three}\\ \text{statements } (\sigma(i)<i),\ (\sigma(i)=i) \text{ and } (\sigma(i)>i)\end{array}\right) \\
&= \underbrace{\sum_{i\in Y} 1}_{=|Y|\cdot 1=|Y|} + \underbrace{\sum_{i\in X} 1}_{=|X|\cdot 1=|X|} + \underbrace{\sum_{i\in Z} 1}_{=|Z|\cdot 1=|Z|} = \underbrace{|Y|}_{=|Z|} + |X| + |Z| \\
&= |Z| + |X| + |Z| = |X| + 2 \cdot |Z| \equiv |X| \bmod 2.
\end{aligned}$$

This proves Lemma 7.170. □

---

[531] *Proof.* Let $j \in Z$. The definition of β yields $\beta(j) = \sigma(j)$. Comparing this with $\underbrace{\sigma^{-1}}_{=\sigma}(j) = \sigma(j)$, we obtain $\beta(j) = \sigma^{-1}(j)$. Now,

$$\begin{aligned}
(\alpha \circ \beta)(j) = \alpha(\beta(j)) &= \sigma\left(\underbrace{\beta(j)}_{=\sigma^{-1}(j)}\right) \qquad (\text{by the definition of } \alpha) \\
&= \sigma\left(\sigma^{-1}(j)\right) = j = \operatorname{id}(j).
\end{aligned}$$

Now, forget that we fixed j. We thus have shown that $(\alpha \circ \beta)(j) = \operatorname{id}(j)$ for each $j \in Z$. In other words, $\alpha \circ \beta = \operatorname{id}$. Qed.

[532] for similar reasons

Corollary 7.171. Let $n \in \mathbb{N}$ be odd. Let $\sigma \in S_n$ be such that $\sigma^{-1} = \sigma$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an alternating $n \times n$-matrix. Then,

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = 0.$$

Proof of Corollary 7.171. Let $X = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) = i\}$. Then, Lemma 7.170 yields $|X| \equiv n \equiv 1 \operatorname{mod} 2$ (since n is odd). If we had $X = \varnothing$, then we would have $\left| \underbrace{X}_{=\varnothing} \right| = |\varnothing| = 0 \not\equiv 1 \operatorname{mod} 2$; this would contradict $|X| \equiv 1 \operatorname{mod} 2$. Hence, we cannot have $X = \varnothing$. Thus, we have $X \neq \varnothing$. In other words, the set X is nonempty. Hence, there exists some $x \in X$. Consider this x.

We have $x \in X = \{i \in \{1, 2, \ldots, n\} \mid \sigma(i) = i\}$. In other words, x is an element i of $\{1, 2, \ldots, n\}$ satisfying $\sigma(i) = i$. In other words, x is an element of $\{1, 2, \ldots, n\}$ and satisfies $\sigma(x) = x$. Now, Lemma 7.169 **(a)** (applied to $i = x$) yields $a_{x,x} = 0$. Since $\sigma(x) = x$, we have $a_{x,\sigma(x)} = a_{x,x} = 0$.

But $x \in \{1, 2, \ldots, n\}$. Hence, $a_{x,\sigma(x)}$ is a factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ (namely, the factor for $i = x$). This factor is 0 (since $a_{x,\sigma(x)} = 0$). Hence, one factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ is 0. Thus, the whole product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ must be 0 (because if one factor of a product is 0, then the whole product must be 0). In other words, $\prod_{i=1}^{n} a_{i,\sigma(i)} = 0$. This proves Corollary 7.171. □

Lemma 7.172. Let $n \in \mathbb{N}$ be odd. Let $\sigma \in S_n$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an alternating $n \times n$-matrix. Then,

$$\prod_{i=1}^{n} a_{i,\sigma^{-1}(i)} = -\prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Proof of Lemma 7.172. We have $\sigma \in S_n$. In other words, σ is a permutation of the set $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$). In other words, σ is a bijection $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$.

Now,

$$\prod_{i=1}^{n} a_{i,\sigma^{-1}(i)} = \prod_{i=1}^{n} \underbrace{a_{\sigma(i),\sigma^{-1}(\sigma(i))}}_{\substack{=a_{\sigma(i),i} \\ (\text{since } \sigma^{-1}(\sigma(i))=i)}}$$
$$\left(\begin{array}{c}\text{here, we have substituted } \sigma(i) \text{ for } i \text{ in the product,} \\ \text{since } \sigma : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\} \text{ is a bijection}\end{array}\right)$$
$$= \prod_{i=1}^{n} \underbrace{a_{\sigma(i),i}}_{\substack{=-a_{i,\sigma(i)} \\ (\text{by Lemma 7.169 } \textbf{(b)} \\ (\text{applied to } (\sigma(i),i) \text{ instead of } (i,j)))}} = \prod_{i=1}^{n} \left(-a_{i,\sigma(i)}\right)$$
$$= \underbrace{(-1)^n}_{\substack{=-1 \\ (\text{since } n \text{ is odd})}} \prod_{i=1}^{n} a_{i,\sigma(i)} = -\prod_{i=1}^{n} a_{i,\sigma(i)}.$$

This proves Lemma 7.172. □

Solution to Exercise 6.24. **(a)** Assume that A is antisymmetric.

Recall that the matrix A is antisymmetric if and only if $A^T = -A$ (by the definition of "antisymmetric"). Hence, $A^T = -A$ (since the matrix A is antisymmetric).

Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Hence,

$$\det A = \det\left(\underbrace{A^T}_{=-A=(-1)A}\right) = \det((-1)A) = \underbrace{(-1)^n}_{\substack{=-1 \\ (\text{since } n \text{ is odd})}} \det A$$
$$(\text{by Proposition 6.12 (applied to } \lambda = -1))$$
$$= (-1)\det A = -\det A.$$

Adding $\det A$ to both sides of this equality, we obtain $\det A + \det A = 0$. Thus, $0 = \det A + \det A = 2\det A$. This solves Exercise 6.24 **(a)**.

(b) Assume that A is alternating.

Write the $n \times n$-matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

The set S_n is finite. Hence, there exists a bijection $\beta : S_n \to \{1,2,\ldots,|S_n|\}$. Consider this bijection β. (Of course, $|S_n| = n!$; but we will not use this.)

Recall that

$$(-1)^{\sigma^{-1}} = (-1)^{\sigma} \qquad \text{for every } \sigma \in S_n. \tag{1195}$$

The map $S_n \to S_n,\ \sigma \mapsto \sigma^{-1}$ is a bijection (indeed, it is its own inverse).

Now,

$$\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)<\beta(\sigma^{-1})}} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}$$

$$= \underbrace{\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma^{-1})<\beta((\sigma^{-1})^{-1})}}}_{\substack{=\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma^{-1})<\beta(\sigma)}}\\ \text{(since } (\sigma^{-1})^{-1}=\sigma\text{)}}} \underbrace{(-1)^{\sigma^{-1}}}_{\substack{=(-1)^{\sigma}\\ \text{(by (1195))}}} \underbrace{\prod_{i=1}^{n} a_{i,\sigma^{-1}(i)}}_{\substack{=-\prod_{i=1}^{n} a_{i,\sigma(i)}\\ \text{(by Lemma 7.172)}}}$$

$$\left(\begin{array}{c}\text{here, we have substituted } \sigma^{-1} \text{ for } \sigma \text{ in the sum,}\\ \text{since the map } S_n \to S_n,\ \sigma \mapsto \sigma^{-1} \text{ is a bijection}\end{array}\right)$$

$$= \underbrace{\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma^{-1})<\beta(\sigma)}}}_{\substack{=\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}}\\ \text{(since for every } \sigma\in S_n\text{, the}\\ \text{statement } (\beta(\sigma^{-1})<\beta(\sigma)) \text{ is}\\ \text{equivalent to } (\beta(\sigma)>\beta(\sigma^{-1})))}} (-1)^{\sigma} \left(-\prod_{i=1}^{n} a_{i,\sigma(i)}\right) = \sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^{\sigma} \left(-\prod_{i=1}^{n} a_{i,\sigma(i)}\right)$$

$$= -\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}. \tag{1196}$$

Also, every $\sigma \in S_n$ satisfying $\beta(\sigma) = \beta(\sigma^{-1})$ must satisfy

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = 0 \tag{1197}$$

[533].

[533] *Proof of (1197):* Let $\sigma \in S_n$ be such that $\beta(\sigma) = \beta(\sigma^{-1})$.

Recall that β is a bijection. Hence, a map β^{-1} is well-defined. Now, $\beta^{-1}\left(\underbrace{\beta(\sigma)}_{=\beta(\sigma^{-1})}\right) = \beta^{-1}(\beta(\sigma^{-1})) = \sigma^{-1}$, so that $\sigma^{-1} = \beta^{-1}(\beta(\sigma)) = \sigma$. Thus, Corollary 7.171 yields $\prod_{i=1}^{n} a_{i,\sigma(i)} = 0$. This proves (1197).

Now, recall that $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence, the definition of $\det A$ yields

$$\begin{aligned}
&\det A \\
&= \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} \\
&= \underbrace{\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)<\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}}_{\substack{= -\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} \\ \text{(by (1196))}}} + \sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)=\beta(\sigma^{-1})}} (-1)^\sigma \underbrace{\prod_{i=1}^n a_{i,\sigma(i)}}_{\substack{=0\\ \text{(by (1197))}}} + \sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} \\
&\qquad \left(\begin{array}{c}\text{since each } \sigma\in S_n \text{ satisfies exactly one of the three}\\ \text{statements } \left(\beta(\sigma)<\beta(\sigma^{-1})\right),\ \left(\beta(\sigma)=\beta(\sigma^{-1})\right) \text{ and } \left(\beta(\sigma)>\beta(\sigma^{-1})\right)\end{array}\right) \\
&= -\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} + \underbrace{\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)=\beta(\sigma^{-1})}} (-1)^\sigma 0}_{=0} + \sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} \\
&= -\sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} + \sum_{\substack{\sigma\in S_n;\\ \beta(\sigma)>\beta(\sigma^{-1})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} = 0.
\end{aligned}$$

This solves Exercise 6.24 **(b)**. □

A short solution of Exercise 6.24 **(b)** using abstract algebra appears in [AndDos12, Theorem 12.2].

7.93. Solution to Exercise 6.25

Solution to Exercise 6.25. Define $n-1$ elements $c_1, c_2, \ldots, c_{n-1}$ of $\mathbb{K}$ by

$$(c_i = 1 \text{ for every } i \in \{1, 2, \ldots, n-1\}).$$

Define an $n\times n$-matrix A as in Definition 6.86. For every two elements x and y of $\{0, 1, \ldots, n\}$ satisfying $x \leq y$, we define a $(y-x)\times(y-x)$-matrix $A_{x,y}$ as in Proposition 6.87.

Now, we claim that

$$u_i = \det(A_{0,i}) \qquad \text{for every } i \in \{0, 1, \ldots, n\}. \tag{1198}$$

[*Proof of (1198):* We shall prove (1198) by strong induction over i. Thus, we fix some $k \in \{0, 1, \ldots, n\}$. We assume that (1198) holds for all $i < k$. We now must show that (1198) holds for $i = k$. In other words, we must show that $u_k = \det(A_{0,k})$.

This holds if $k \leq 1$ [534]. Hence, we can WLOG assume that we don't have $k \leq 1$. Assume this.

We have $k > 1$ (since we don't have $k \leq 1$). Thus, $k \geq 2$. Hence, $k-1$ and $k-2$ are nonnegative integers satisfying $k-1 < k$ and $k-2 < k$. Hence, we can apply (1198) to $i = k-1$ (since we have assumed that (1198) holds for all $i < k$). As a result, we obtain $u_{k-1} = \det(A_{0,k-1})$. Also, we can apply (1198) to $i = k-2$ (since we have assumed that (1198) holds for all $i < k$). As a result, we obtain $u_{k-2} = \det(A_{0,k-2})$.

We have $c_{k-1} = 1$ (by the definition of c_{k-1}).

Now, $0 \leq k-2$ (since $k \geq 2$). Hence, Proposition 6.87 **(c)** (applied to $x = 0$ and $y = k$) yields

$$\det(A_{0,k}) = a_k \det(A_{0,k-1}) - b_{k-1} \underbrace{c_{k-1}}_{=1} \det(A_{0,k-2}) = a_k \det(A_{0,k-1}) - b_{k-1} \det(A_{0,k-2}).$$

Comparing this with

$$u_k = a_k \underbrace{u_{k-1}}_{=\det(A_{0,k-1})} - b_{k-1} \underbrace{u_{k-2}}_{=\det(A_{0,k-2})} \qquad \text{(by the recursive definition of } (u_0, u_1, \ldots, u_n)\text{)}$$
$$= a_k \det(A_{0,k-1}) - b_{k-1} \det(A_{0,k-2}),$$

we obtain $u_k = \det(A_{0,k})$. In other words, (1198) holds for $i = k$. This completes the induction step. Thus, (1198) is proven.]

Next, we claim that

$$v_i = \det(A_{n-i,n}) \qquad \text{for every } i \in \{0, 1, \ldots, n\}. \qquad (1199)$$

[*Proof of (1199):* We shall prove (1199) by strong induction over i. Thus, we fix some $k \in \{0, 1, \ldots, n\}$. We assume that (1199) holds for all $i < k$. We now must show that (1199) holds for $i = k$. In other words, we must show that $v_k = \det(A_{n-k,n})$.

This holds if $k \leq 1$ [535]. Hence, we can WLOG assume that we don't have $k \leq 1$.

---

[534] *Proof.* Proposition 6.87 **(a)** (applied to $x = 0$) yields $\det(A_{0,0}) = 1$. Compared with $u_0 = 1$, this yields $u_0 = \det(A_{0,0})$.

Proposition 6.87 **(b)** (applied to $x = 0$) yields $\det(A_{0,1}) = a_1$. Compared with $u_1 = a_1$, this yields $u_1 = \det(A_{0,1})$.

Now, $u_k = \det(A_{0,k})$ holds if $k = 0$ (because we have $u_0 = \det(A_{0,0})$), and also holds if $k = 1$ (since $u_1 = \det(A_{0,1})$). Therefore, $u_k = \det(A_{0,k})$ holds if $k \leq 1$. Qed.

[535] *Proof.* Proposition 6.87 **(a)** (applied to $x = n$) yields $\det(A_{n,n}) = 1$. Thus, $\det(A_{n-0,n}) = \det(A_{n,n}) = 1$. Compared with $v_0 = 1$, this yields $v_0 = \det(A_{n-0,n})$.

Proposition 6.87 **(b)** (applied to $x = n-1$) yields $\det\left(A_{n-1,(n-1)+1}\right) = a_{(n-1)+1}$. This rewrites as $\det(A_{n-1,n}) = a_n$ (since $(n-1)+1 = n$). Compared with $v_1 = a_n$, this yields $v_1 = \det(A_{n-1,n})$.

Now, $v_k = \det(A_{n-k,n})$ holds if $k = 0$ (because we have $v_0 = \det(A_{n-0,n})$), and also holds if $k = 1$ (since $v_1 = \det(A_{n-1,n})$). Therefore, $v_k = \det(A_{n-k,n})$ holds if $k \leq 1$. Qed.

Assume this.

We have $k > 1$ (since we don't have $k \leq 1$). Thus, $k \geq 2$. Hence, $k-1$ and $k-2$ are nonnegative integers satisfying $k-1 < k$ and $k-2 < k$. Hence, we can apply (1199) to $i = k-1$ (since we have assumed that (1199) holds for all $i < k$). As a result, we obtain $v_{k-1} = \det\left(\underbrace{A_{n-(k-1),n}}_{=A_{n-k+1,n}}\right) = \det\left(A_{n-k+1,n}\right)$. Also, we can apply (1199) to $i = k-2$ (since we have assumed that (1199) holds for all $i < k$). As a result, we obtain $v_{k-2} = \det\left(\underbrace{A_{n-(k-2),n}}_{=A_{n-k+2,n}}\right) = \det\left(A_{n-k+2,n}\right)$.

We have $c_{n-k+1} = 1$ (by the definition of c_{n-k+1}).

Now, $n-k \leq n-2$ (since $k \geq 2$). Hence, Proposition 6.87 **(d)** (applied to $x = n-k$ and $y = n$) yields

$$\begin{aligned}\det\left(A_{n-k,n}\right) &= a_{n-k+1}\det\left(A_{n-k+1,n}\right) - b_{n-k+1}\underbrace{c_{n-k+1}}_{=1}\det\left(A_{n-k+2,n}\right) \\ &= a_{n-k+1}\det\left(A_{n-k+1,n}\right) - b_{n-k+1}\det\left(A_{n-k+2,n}\right).\end{aligned}$$

Comparing this with

$$\begin{aligned}v_k &= a_{n-k+1}\underbrace{v_{k-1}}_{=\det\left(A_{n-k+1,n}\right)} - b_{n-k+1}\underbrace{v_{k-2}}_{=\det\left(A_{n-k+2,n}\right)} \\ &\qquad \left(\text{by the recursive definition of } \left(v_0, v_1, \ldots, v_n\right)\right) \\ &= a_{n-k+1}\det\left(A_{n-k+1,n}\right) - b_{n-k+1}\det\left(A_{n-k+2,n}\right),\end{aligned}$$

we obtain $v_k = \det\left(A_{n-k,n}\right)$. In other words, (1199) holds for $i = k$. This completes the induction step. Thus, (1199) is proven.]

Now, we are almost done. In fact, applying (1198) to $i = n$, we obtain $u_n = \det\left(A_{0,n}\right)$. On the other hand, applying (1199) to $i = n$, we obtain $v_n = \det\left(A_{n-n,n}\right) = \det\left(A_{0,n}\right)$ (since $n - n = 0$). Comparing this with $u_n = \det\left(A_{0,n}\right)$, we obtain $u_n = v_n$. Exercise 6.25 is solved. $\square$

7.94. Solution to Exercise 6.26

Solution to Exercise 6.26. Assume that all denominators appearing in Exercise 6.26 are invertible. For every $k \in \{1, 2, \ldots, n\}$, define an element p_k of $\mathbb{K}$ by

$$p_k = a_k - \cfrac{b_k c_k}{a_{k+1} - \cfrac{b_{k+1} c_{k+1}}{a_{k+2} - \cfrac{b_{k+2} c_{k+2}}{a_{k+3} - \cfrac{\ddots}{- \cfrac{b_{n-2} c_{n-2}}{a_{n-1} - \cfrac{b_{n-1} c_{n-1}}{a_n}}}}}}$$

[536]. This definition of p_k immediately gives a recursion:

- We have $p_n = a_n$.
- For every $k \in \{1, 2, \ldots, n-1\}$, we have
$$p_k = a_k - \frac{b_k c_k}{p_{k+1}}. \tag{1200}$$

Now, we shall show that

$$\det(A_{n-k-1,n}) = p_{n-k} \det(A_{n-k,n}) \qquad \text{for every } k \in \{0, 1, \ldots, n-1\}. \tag{1201}$$

[*Proof of (1201):* We shall prove (1201) by induction over k:

Induction base: It is easy to see that (1201) holds for $k = 0$ [537]. This completes the induction base.

Induction step: Let $K \in \{0, 1, \ldots, n-1\}$ be positive. Assume that (1201) holds for $k = K - 1$. We shall show that (1201) holds for $k = K$.

[536] If $k = n$, then this should be interpreted as saying that $p_n = a_n$.

[537] *Proof.* Proposition 6.87 **(a)** (applied to $x = n$) yields $\det(A_{n,n}) = 1$. Hence, $\underbrace{p_{n-0}}_{=p_n=a_n} \det\left(\underbrace{A_{n-0,n}}_{=A_{n,n}}\right) = a_n$.

Proposition 6.87 **(b)** (applied to $x = n-1$) yields $\det\left(A_{n-1,(n-1)+1}\right) = a_{(n-1)+1}$. This rewrites as $\det(A_{n-1,n}) = a_n$ (since $(n-1)+1 = n$). Thus, $\det\left(\underbrace{A_{n-0-1,n}}_{=A_{n-1,n}}\right) = \det(A_{n-1,n}) = a_n$. Comparing this with $\underbrace{p_{n-0}}_{=p_n=a_n} \det\left(\underbrace{A_{n-0,n}}_{=A_{n,n}}\right) = a_n \underbrace{\det(A_{n,n})}_{=1} = a_n$, we obtain $\det(A_{n-0-1,n}) = p_{n-0} \det(A_{n-0,n})$. In other words, (1201) holds for $k = 0$. Qed.

We have assumed that (1201) holds for $k = K - 1$. In other words, we have $\det\left(A_{n-(K-1)-1,n}\right) = p_{n-(K-1)} \det\left(A_{n-(K-1),n}\right)$. Since $n - (K-1) = n - K + 1$, this rewrites as $\det(A_{n-K+1-1,n}) = p_{n-K+1} \det(A_{n-K+1,n})$. Solving this for $\det(A_{n-K+1,n})$, we obtain

$$\det(A_{n-K+1,n}) = \frac{\det(A_{n-K+1-1,n})}{p_{n-K+1}} = \frac{1}{p_{n-K+1}} \det\left(\underbrace{A_{n-K+1-1,n}}_{=A_{n-K,n}}\right) = \frac{1}{p_{n-K+1}} \det(A_{n-K,n}). \tag{1202}$$

We have $K \in \{1, 2, \ldots, n-1\}$ (since K is positive and belongs to $\{0, 1, \ldots, n-1\}$). Hence, applying (1200) to $k = n - K$, we obtain

$$p_{n-K} = a_{n-K} - \frac{b_{n-K}c_{n-K}}{p_{n-K+1}}. \tag{1203}$$

We have $K \leq n-1$ and thus $n - K - 1 \geq 0$. Moreover, $K \geq 1$ (since K is positive), thus $n - \underbrace{K}_{\geq 1} - 1 \leq n - 1 - 1 = n - 2$. Hence, Proposition 6.87 **(d)** (applied to $x = n - K - 1$ and $y = n$) shows that

$$\begin{aligned}
&\det(A_{n-K-1,n}) \\
&= \underbrace{a_{(n-K-1)+1}}_{=a_{n-K}} \det\left(\underbrace{A_{(n-K-1)+1,n}}_{=A_{n-K,n}}\right) - \underbrace{b_{(n-K-1)+1}}_{=b_{n-K}} \underbrace{c_{(n-K-1)+1}}_{=c_{n-K}} \det\left(\underbrace{A_{(n-K-1)+2,n}}_{=A_{n-K+1,n}}\right) \\
&= a_{n-K} \det(A_{n-K,n}) - b_{n-K}c_{n-K} \underbrace{\det(A_{n-K+1,n})}_{\substack{=\frac{1}{p_{n-K+1}}\det(A_{n-K,n}) \\ \text{(by (1202))}}} \\
&= a_{n-K} \det(A_{n-K,n}) - b_{n-K}c_{n-K} \cdot \frac{1}{p_{n-K+1}} \det(A_{n-K,n}) \\
&= \underbrace{\left(a_{n-K} - b_{n-K}c_{n-K} \cdot \frac{1}{p_{n-K+1}}\right)}_{\substack{=a_{n-K} - \frac{b_{n-K}c_{n-K}}{p_{n-K+1}} = p_{n-K} \\ \text{(by (1203))}}} \det(A_{n-K,n}) = p_{n-K} \det(A_{n-K,n}).
\end{aligned}$$

In other words, (1201) holds for $k = K$. This completes the induction step. Thus, (1201) is proven by induction.]

Now, we can apply (1201) to $k = n-1$. This gives us

$$\det\left(A_{n-(n-1)-1,n}\right) = \underbrace{p_{n-(n-1)}}_{=p_1} \det\left(\underbrace{A_{n-(n-1),n}}_{=A_{1,n}}\right) = p_1 \det\left(A_{1,n}\right).$$

Since $n-(n-1)-1=0$, this rewrites as $\det\left(A_{0,n}\right) = p_1 \det\left(A_{1,n}\right)$. Since $A_{0,n} = A$ (by Proposition 6.87 **(e)**), this simplifies to $\det A = p_1 \det\left(A_{1,n}\right)$. Hence,

$$\frac{\det A}{\det\left(A_{1,n}\right)} = p_1 = a_1 - \cfrac{b_1c_1}{a_2 - \cfrac{b_2c_2}{a_3 - \cfrac{b_3c_3}{a_4 - \ddots - \cfrac{b_{n-2}c_{n-2}}{a_{n-1} - \cfrac{b_{n-1}c_{n-1}}{a_n}}}}}$$

(by the definition of p_1). □

7.95. Solution to Exercise 6.27

First solution to Exercise 6.27. We claim that

$$\det\left(A_{0,i}\right) = f_{i+1} \qquad \text{for every } i \in \{0,1,\ldots,n\}. \tag{1204}$$

[*Proof of (1204):* We shall prove (1204) by strong induction over i. Thus, we fix some $k \in \{0,1,\ldots,n\}$. We assume that (1204) holds for all $i < k$. We now must show that (1204) holds for $i = k$. In other words, we must show that $\det\left(A_{0,k}\right) = f_{k+1}$.

This holds if $k \leq 1$ [538]. Hence, we can WLOG assume that we don't have $k \leq 1$. Assume this.

We have $k > 1$ (since we don't have $k \leq 1$). Thus, $k \geq 2$. Hence, $k-1$ and $k-2$ are nonnegative integers satisfying $k-1 < k$ and $k-2 < k$. Hence, we can apply (1204) to $i = k-1$ (since we have assumed that (1204) holds for all $i < k$). As a result, we obtain $\det\left(A_{0,k-1}\right) = f_{(k-1)+1}$. Also, we can apply (1204) to $i = k-2$ (since we have assumed that (1204) holds for all $i < k$). As a result, we obtain $\det\left(A_{0,k-2}\right) = f_{(k-2)+1}$.

[538] *Proof.* Proposition 6.87 **(a)** (applied to $x = 0$) yields $\det\left(A_{0,0}\right) = 1$. Compared with $f_{0+1} = f_1 = 1$, this yields $\det\left(A_{0,0}\right) = f_{0+1}$.

Proposition 6.87 **(b)** (applied to $x = 0$) yields $\det\left(A_{0,1}\right) = a_1 = 1$. Compared with $f_{1+1} = f_2 = 1$, this yields $\det\left(A_{0,1}\right) = f_{1+1}$.

Now, $\det\left(A_{0,k}\right) = f_{k+1}$ holds if $k = 0$ (because we have $\det\left(A_{0,0}\right) = f_{0+1}$), and also holds if $k = 1$ (since $\det\left(A_{0,1}\right) = f_{1+1}$). Therefore, $\det\left(A_{0,k}\right) = f_{k+1}$ holds if $k \leq 1$. Qed.

Now, $0 \leq k-2$ (since $k \geq 2$). Hence, Proposition 6.87 **(c)** (applied to $x=0$ and $y=k$) yields

$$\det\left(A_{0,k}\right) = \underbrace{a_k}_{\substack{=1 \\ \text{(by the} \\ \text{definition of } a_k)}} \underbrace{\det\left(A_{0,k-1}\right)}_{\substack{=f_{(k-1)+1} \\ =f_k}} - \underbrace{b_{k-1}}_{\substack{=1 \\ \text{(by the} \\ \text{definition of } b_{k-1})}} \underbrace{c_{k-1}}_{\substack{=-1 \\ \text{(by the} \\ \text{definition of } c_{k-1})}} \underbrace{\det\left(A_{0,k-2}\right)}_{\substack{=f_{(k-2)+1} \\ =f_{k-1}}}$$
$$= f_k - (-1) f_{k-1} = f_k + f_{k-1}.$$

Comparing this with

$$f_{k+1} = f_k + f_{k-1} \qquad \text{(by the recursive definition of the Fibonacci numbers)},$$

we obtain $\det\left(A_{0,k}\right) = f_{k+1}$. In other words, (1204) holds for $i=k$. This completes the induction step. Thus, (1204) is proven.]

Now, applying (1204) to $i=n$, we obtain $\det\left(A_{0,n}\right) = f_{n+1}$. Since $A_{0,n} = A$ (by Proposition 6.87 **(e)**), we can rewrite this as $\det A = f_{n+1}$. This solves Exercise 6.27. □

Second solution to Exercise 6.27 (sketched). Here is a different solution for Exercise 6.27, which is far more complicated than the previous one, but has the pedagogical advantage of illuminating the connection between determinants and permutations, and the combinatorics of the latter.

Exercise 4.3 (applied to $n+1$ instead of n) shows that f_{n+1} is the number of subsets I of $\{1,2,\ldots,n-1\}$ such that no two elements of I are consecutive. We shall refer to such subsets I as *lacunar sets*.[539] Thus, f_{n+1} is the number of all lacunar sets.

(For example, if $n=5$, then the lacunar sets are $\varnothing$, $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{1,3\}$, $\{1,4\}$, and $\{2,4\}$. Their number, unsurprisingly, is $8 = f_{5+1}$.)

For any $\sigma \in S_n$, we define the following terminology:

- The *excedances* of σ are the elements $i \in \{1,2,\ldots,n\}$ satisfying $\sigma(i) > i$. For instance, the permutation in S_7 written in one-line notation as $(3,1,2,4,5,7,6)$ has excedances 1 and 6.

- We let $\operatorname{Exced}\sigma$ denote the set of all excedances of σ.

- A permutation $\sigma \in S_n$ is said to be *short-legged* if every $i \in \{1,2,\ldots,n\}$ satisfies $|\sigma(i) - i| \leq 1$. For instance, the permutation in S_7 written in one-line notation as $(1,2,4,3,5,7,6)$ is short-legged.

(Most of the terminology here is my own, tailored for this exercise; only the notion of "excedance" is standard. I chose the name "short-legged" because a permutation σ satisfying $|\sigma(i) - i| \leq 1$ "does not take i very far".)

[539] We keep n fixed, so a "lacunar set" will always be a subset of $\{1,2,\ldots,n-1\}$.

What does this all have to do with the exercise? Let us write our matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Thus, for every $(i,j) \in \{1,2,\ldots,n\}^2$, we have

$$a_{i,j} = \begin{cases} a_i, & \text{if } i = j; \\ b_i, & \text{if } i = j-1; \\ c_j, & \text{if } i = j+1; \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 1, & \text{if } i = j; \\ 1, & \text{if } i = j-1; \\ -1, & \text{if } i = j+1; \\ 0, & \text{otherwise} \end{cases} \qquad (1205)$$

(since $a_i = 1$, $b_i = 1$ and $c_j = -1$). Notice that, as a consequence of this equality, we have

$$a_{i,j} = 0 \qquad \text{whenever } |i-j| > 1. \qquad (1206)$$

Now, recall that $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, so that $A^T = \left(a_{j,i}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. But Exercise 6.4 yields $\det\left(A^T\right) = \det A$. Hence,

$$\det A = \det\left(A^T\right) = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n a_{\sigma(i),i} \qquad (1207)$$

(by (341), applied to A^T and $a_{j,i}$ instead of A and $a_{i,j}$). This is an expression for $\det A$, but in order to get any mileage out of it we need to simplify the terms $\prod_{i=1}^n a_{\sigma(i),i}$ for $\sigma \in S_n$. This turns out to depend on whether the permutation σ is short-legged or not:

- If a permutation $\sigma \in S_n$ is not short-legged, then

$$\prod_{i=1}^n a_{\sigma(i),i} = 0. \qquad (1208)$$

[*Proof of (1208):* Let $\sigma \in S_n$ be not short-legged. Thus, there exists a $k \in \{1,2,\ldots,n\}$ satisfying $|\sigma(k) - k| > 1$. The factor $a_{\sigma(k),k}$ corresponding to this k must be 0 (because of (1206)); this forces the whole product $\prod_{i=1}^n a_{\sigma(i),i}$ to become 0. Thus, (1208) follows.]

- If a permutation $\sigma \in S_n$ is short-legged, then

$$\prod_{i=1}^n a_{\sigma(i),i} = (-1)^{|\operatorname{Exced}\sigma|}. \qquad (1209)$$

[*Proof of (1209):* Let $\sigma \in S_n$ be short-legged. Thus, every $i \in \{1,2,\ldots,n\}$

satisfies $|\sigma(i) - i| \leq 1$. Consequently, for every $i \in \{1, 2, \ldots, n\}$, we have

$$a_{\sigma(i),i} = \begin{cases} 1, & \text{if } \sigma(i) = i; \\ 1, & \text{if } \sigma(i) = i-1; \\ -1, & \text{if } \sigma(i) = i+1; \\ 0, & \text{otherwise} \end{cases} \qquad \text{(by (1205))}$$
$$= \begin{cases} 1, & \text{if } \sigma(i) = i; \\ 1, & \text{if } \sigma(i) = i-1; \\ -1, & \text{if } \sigma(i) = i+1 \end{cases}$$
$$\left(\begin{array}{c}\text{since the inequality } |\sigma(i) - i| \leq 1 \text{ ensures that one of the} \\ \text{conditions } \sigma(i) = i, \ \sigma(i) = i-1 \text{ and } \sigma(i) = i+1 \text{ must hold}\end{array}\right)$$
$$= \begin{cases} 1, & \text{if } \sigma(i) \leq i; \\ -1, & \text{if } \sigma(i) > i \end{cases}.$$

Thus,

$$\prod_{i=1}^{n} a_{\sigma(i),i} = \prod_{i=1}^{n} \begin{cases} 1, & \text{if } \sigma(i) \leq i; \\ -1, & \text{if } \sigma(i) > i \end{cases} = (-1)^{(\text{the number of all } i \in \{1,2,\ldots,n\} \text{ satisfying } \sigma(i) > i)}$$
$$= (-1)^{|\{i \in \{1,2,\ldots,n\} \ | \ \sigma(i) > i\}|} = (-1)^{|\operatorname{Exced} \sigma|};$$

thus, (1209) is proven.]

Now, (1207) becomes

$$\det A = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} a_{\sigma(i),i}$$
$$= \sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is not short-legged}}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} a_{\sigma(i),i}}_{\substack{=0 \\ \text{(by (1208))}}} + \sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is short-legged}}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} a_{\sigma(i),i}}_{\substack{=(-1)^{|\operatorname{Exced} \sigma|} \\ \text{(by (1209))}}}$$
$$= \underbrace{\sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is not short-legged}}} (-1)^{\sigma} 0}_{=0} + \sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is short-legged}}} (-1)^{\sigma} (-1)^{|\operatorname{Exced} \sigma|}$$
$$= \sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is short-legged}}} (-1)^{\sigma} (-1)^{|\operatorname{Exced} \sigma|}. \tag{1210}$$

We still don't see how this connects to f_{n+1}, though. So let us relate short-legged permutations to lacunar sets.

For any lacunar set I, we can define a permutation $\tau_I \in S_n$ by the following rule:

$$\tau_I(k) = \begin{cases} k+1, & \text{if } k \in I; \\ k-1, & \text{if } k-1 \in I; \\ k, & \text{otherwise} \end{cases} \qquad \text{for every } k \in \{1,2,\ldots,n\}.$$

In other words, τ_I is the permutation of $\{1,2,\ldots,n\}$ which interchanges every element i of I with its successor $i+1$, while leaving all remaining elements unchanged. Make sure you understand why τ_I is a well-defined map $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ [540] and a permutation[541].

(For example, if $n = 7$ and $I = \{2,5\}$, then $\tau_I = (1,3,2,4,6,5,7)$ in one-line notation.)

It is clear that the permutation τ_I is short-legged. Moreover, it satisfies

$$\operatorname{Exced}(\tau_I) = I; \tag{1211}$$

as a consequence, it is possible to reconstruct I from τ_I. Thus, the permutations τ_I for distinct I are distinct.

What is the sign $(-1)^{\tau_I}$? It is easy to see (from the construction of τ_I) that the only inversions of τ_I are the pairs $(i, i+1)$ for $i \in I$ (essentially, the short-leggedness of τ_I prevents τ_I from changing the order of two non-adjacent integers). Thus, the number of these inversions is $|I|$. Thus, $\ell(\tau_I) = |I|$. Hence,

$$(-1)^{\tau_I} = (-1)^{\ell(\tau_I)} = (-1)^{|I|}. \tag{1212}$$

So there are at least some short-legged permutations that we understand well: the τ_I for lacunar sets I. Are there others?

It turns out that there aren't. Indeed,

$$\text{every short-legged } \sigma \in S_n \text{ has the form } \tau_I \text{ for some lacunar set } I. \tag{1213}$$

Before we can prove this, we shall prove two auxiliary observations:

[540] Here are the two things you need to check:

- The term $\begin{cases} k+1, & \text{if } k \in I; \\ k-1, & \text{if } k-1 \in I; \\ k, & \text{otherwise} \end{cases}$ is unambiguous, because no k satisfies both $k \in I$ and $k-1 \in I$ at the same time. (Here is where you use the lacunarity of I.)
- We have $\begin{cases} k+1, & \text{if } k \in I; \\ k-1, & \text{if } k-1 \in I; \\ k, & \text{otherwise} \end{cases} \in \{1,2,\ldots,n\}$ for every $k \in \{1,2,\ldots,n\}$. (Here you use $I \subseteq \{1,2,\ldots,n-1\}$. If I were only a subset of $\{1,2,\ldots,n\}$, then this would fall outside of $\{1,2,\ldots,n\}$ for $k = n$.)

[541] Indeed, $\tau_I \circ \tau_I = \operatorname{id}$, so that τ_I is its own inverse.

Observation 1: Let $\sigma \in S_n$ be short-legged. If $i \in \{1, 2, \ldots, n\}$ be such that $\sigma(i) = i + 1$, then $\sigma(i+1) = i$.

Observation 2: Let $\sigma \in S_n$ be short-legged. If $i \in \{1, 2, \ldots, n\}$ be such that $\sigma(i) = i - 1$, then $\sigma(i-1) = i$.

[*Proof of Observation 1.* Assume the contrary. Thus, there exists some $i \in \{1, 2, \ldots, n\}$ such that $\sigma(i) = i + 1$ but $\sigma(i+1) \neq i$. We call such i's *evil*. By our assumption, there exists at least one evil i. Consider the highest evil i. Thus, $i + 1$ is not evil.

Since i is evil, we have $\sigma(i) = i + 1$ but $\sigma(i+1) \neq i$. In particular, $i + 1 = \sigma(i) \in \{1, 2, \ldots, n\}$, so that $\sigma(i+1)$ is well-defined. Since σ is short-legged, we have $|\sigma(i+1) - (i+1)| \leq 1$. Hence, $\sigma(i+1)$ is either i or $i+1$ or $i+2$. But $\sigma(i+1)$ cannot be i (since $\sigma(i+1) \neq i$) and cannot be $i+1$ either (since this would cause $\sigma(i+1) = i + 1 = \sigma(i)$, which would contradict the injectivity of σ). Hence, $\sigma(i+1)$ must be $i+2$. In other words, $\sigma(i+1) = i+2$. Moreover, the injectivity of σ shows that $\sigma(i+2) \neq \sigma(i) = i+1$, so that $i+1$ is evil. But this contradicts the fact that $i+1$ is not evil. Thus, Observation 1 is proven.]

[*Proof of Observation 2.* Analogous to Observation 1 (this time, take the lowest evil i), and left to the reader.]

[*Proof of (1213):* Let $\sigma \in S_n$ be short-legged. We must show that $\sigma = \tau_I$ for some lacunar set I.

We set $I = \operatorname{Exced} \sigma$. (This is the only choice we can make to have any hope for $\sigma = \tau_I$ to be true; indeed, (1211) ensures that if $\sigma = \tau_I$, then $\operatorname{Exced} \sigma = I$.)

We notice that

$$\sigma(i) = i + 1 \qquad \text{for every } i \in I \tag{1214}$$

[542]. Thus, n cannot belong to I (since this would entail $\sigma(n) = n + 1$, but $n + 1 \notin \{1, 2, \ldots, n\}$). Hence, $I \subseteq \{1, 2, \ldots, n-1\}$.

Let us first show that I is a lacunar set. Indeed, assume (for the sake of contradiction) that this is not so. Then, there exists some $i \in I$ such that $i + 1 \in I$. Consider such an i. We have $\sigma(i) = i + 1$ (by (1214)), and thus $\sigma(i+1) = i$ (by Observation 1). But $\sigma(i+1) = i + 2$ (by (1214), applied to $i+1$ instead of i). Hence, $i + 2 = \sigma(i+1) = i$, which is absurd. Hence, we have found a contradiction. This finishes our proof that I is a lacunar set.

We still need to show that we actually have $\sigma = \tau_I$. In other words, we need to show that $\sigma(k) = \tau_I(k)$ for every $k \in \{1, 2, \ldots, n\}$.

So let us fix $k \in \{1, 2, \ldots, n\}$, and let us show that $\sigma(k) = \tau_I(k)$. We are in one of the following three cases:

Case 1: We have $k \in I$.

Case 2: We have $k - 1 \in I$.

Case 3: Neither $k \in I$ nor $k - 1 \in I$.

- Let us first consider Case 1. In this case, $k \in I$. Hence, (1214) (applied to $i = k$) yields $\sigma(k) = k + 1$. On the other hand, the definition of τ_I shows that

[542] *Proof of (1214):* Let $i \in I$. Thus, $i \in I = \operatorname{Exced} \sigma$. In other words, i is an excedance of σ. Hence, $\sigma(i) > i$. Since $|\sigma(i) - i| \leq 1$ (because σ is short-legged), this means that $\sigma(i) = i + 1$, qed.

$\tau_I(k) = k+1$ as well. Thus, $\sigma(k) = k+1 = \tau_I(k)$. Hence, $\sigma(k) = \tau_I(k)$ is proven in Case 1.

- Let us now consider Case 2. In this case, $k-1 \in I$. Hence, (1214) (applied to $i = k-1$) yields $\sigma(k-1) = k$. Since σ is injective, we have $\sigma(k) \neq \sigma(k-1) = k$. Also, I is lacunar, so that $k-1 \in I$ entails $k \notin I$; thus, $k \notin I = \operatorname{Exced}\sigma$, so that k is not an excedance of σ. In other words, $\sigma(k) \leq k$. Combined with $\sigma(k) \neq k$, this yields $\sigma(k) < k$. Since $|\sigma(k) - k| \leq 1$ (because σ is short-legged), this shows that $\sigma(k) = k-1$. On the other hand, $\tau_I(k) = k-1$ by the definition of τ_I. Thus, $\sigma(k) = k-1 = \tau_I(k)$. Hence, $\sigma(k) = \tau_I(k)$ is proven in Case 2.

- Let us finally consider Case 3. In this case, neither $k \in I$ nor $k-1 \in I$. Let us now show that $\sigma(k) = k$. Indeed, assume the contrary. Thus, $\sigma(k) \neq k$. As in Case 2, we can use this (and $k \notin I$) to show that $\sigma(k) = k-1$. Observation 2 thus shows that $\sigma(k-1) = k > k-1$, so that $k-1$ is an excedance of σ. In other words, $k-1 \in \operatorname{Exced}\sigma = I$. This contradicts the assumption that we do not have $k-1 \in I$. This contradiction concludes our proof of $\sigma(k) = k$. On the other hand, $\tau_I(k) = k$ by the definition of τ_I. Thus, $\sigma(k) = k = \tau_I(k)$. Hence, $\sigma(k) = \tau_I(k)$ is proven in Case 3.

We now have shown that $\sigma(k) = \tau_I(k)$ in all possible cases. Thus, $\sigma = \tau_I$. Since I is a lacunar set, this proves (1213).]

All we now need to do is combine our results. For every lacunar set I, we have defined a short-legged permutation $\tau_I \in S_n$. Conversely, we know (from (1213)) that every short-legged $\sigma \in S_n$ has the form τ_I for some lacunar set I; we also know that this I is uniquely determined by the σ (since the permutations τ_I for distinct I are distinct). Thus, we have a bijection between the lacunar sets and the short-legged permutations in S_n; the bijection sends every I to τ_I. Consequently,

$$\begin{aligned}
&\sum_{\substack{\sigma \in S_n; \\ \sigma \text{ is short-legged}}} (-1)^{\sigma} (-1)^{|\operatorname{Exced}\sigma|} \\
&= \sum_{I \text{ is a lacunar set}} \underbrace{(-1)^{\tau_I}}_{\substack{=(-1)^{|I|} \\ \text{(by (1212))}}} \underbrace{(-1)^{|\operatorname{Exced}(\tau_I)|}}_{\substack{=(-1)^{|I|} \\ \text{(by (1211))}}} \\
&= \sum_{I \text{ is a lacunar set}} \underbrace{(-1)^{|I|} (-1)^{|I|}}_{=\left((-1)^{|I|}\right)^2 = 1} = \sum_{I \text{ is a lacunar set}} 1 \\
&= (\text{the number of all lacunar sets}) = f_{n+1}.
\end{aligned}$$

Combining this with (1210), we conclude that $\det A = f_{n+1}$. □

7.96. Solution to Exercise 6.28

Solution to Exercise 6.28. This is another case where the solution is really clear with the appropriate amount of waving hands and pointing fingers, but on paper becomes nearly impossible to convey. I shall therefore resort to formalism and computation.

Write the matrices A, B, C, D, A', B', C' and D' in the forms

$$\begin{aligned} A &= \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}, & B &= \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m'}, \\ C &= \left(c_{i,j}\right)_{1\leq i\leq n',\ 1\leq j\leq m}, & D &= \left(d_{i,j}\right)_{1\leq i\leq n',\ 1\leq j\leq m'}, \\ A' &= \left(a'_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq \ell}, & B' &= \left(b'_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq \ell'}, \\ C' &= \left(c'_{i,j}\right)_{1\leq i\leq m',\ 1\leq j\leq \ell}, & D' &= \left(d'_{i,j}\right)_{1\leq i\leq m',\ 1\leq j\leq \ell'}. \end{aligned}$$

The definition of the $(n+n')\times(m+m')$-matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ shows that[543]

$$\begin{aligned} \begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} & b_{1,1} & b_{1,2} & \cdots & b_{1,m'} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} & b_{2,1} & b_{2,2} & \cdots & b_{2,m'} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} & b_{n,1} & b_{n,2} & \cdots & b_{n,m'} \\ c_{1,1} & c_{1,2} & \cdots & c_{1,m} & d_{1,1} & d_{1,2} & \cdots & d_{1,m'} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} & d_{2,1} & d_{2,2} & \cdots & d_{2,m'} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n',1} & c_{n',2} & \cdots & c_{n',m} & d_{n',1} & d_{n',2} & \cdots & d_{n',m'} \end{pmatrix} \\ &= \left(\begin{cases} a_{i,j}, & \text{if } i\leq n \ \&\ j\leq m; \\ b_{i,j-m}, & \text{if } i\leq n \ \&\ j> m; \\ c_{i-n,j}, & \text{if } i> n \ \&\ j\leq m; \\ d_{i-n,j-m}, & \text{if } i> n \ \&\ j> m \end{cases} \right)_{1\leq i\leq n+n',\ 1\leq j\leq m+m'} . \end{aligned}$$

[543]Here and in the following, we use the symbol "&" as shorthand for the word "and".

Similarly,

$$\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} a'_{1,1} & a'_{1,2} & \cdots & a'_{1,\ell} & b'_{1,1} & b'_{1,2} & \cdots & b'_{1,\ell'} \\ a'_{2,1} & a'_{2,2} & \cdots & a'_{2,\ell} & b'_{2,1} & b'_{2,2} & \cdots & b'_{2,\ell'} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a'_{m,1} & a'_{m,2} & \cdots & a'_{m,\ell} & b'_{m,1} & b'_{m,2} & \cdots & b'_{m,\ell'} \\ c'_{1,1} & c'_{1,2} & \cdots & c'_{1,\ell} & d'_{1,1} & d'_{1,2} & \cdots & d'_{1,\ell'} \\ c'_{2,1} & c'_{2,2} & \cdots & c'_{2,\ell} & d'_{2,1} & d'_{2,2} & \cdots & d'_{2,\ell'} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c'_{m',1} & c'_{m',2} & \cdots & c'_{m',\ell} & d'_{m',1} & d'_{m',2} & \cdots & d'_{m',\ell'} \end{pmatrix}$$
$$= \left(\begin{cases} a'_{i,j}, & \text{if } i \leq m \ \& \ j \leq \ell; \\ b'_{i,j-\ell}, & \text{if } i \leq m \ \& \ j > \ell; \\ c'_{i-m,j}, & \text{if } i > m \ \& \ j \leq \ell; \\ d'_{i-m,j-\ell}, & \text{if } i > m \ \& \ j > \ell \end{cases} \right)_{1\leq i\leq m+m',\ 1\leq j\leq \ell+\ell'} .$$

Using these two equalities, we can compute the product $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$: Namely,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$$
$$= \left(\sum_{k=1}^{m+m'} \begin{cases} a_{i,k}, & \text{if } i \leq n \ \& \ k \leq m; \\ b_{i,k-m}, & \text{if } i \leq n \ \& \ k > m; \\ c_{i-n,k}, & \text{if } i > n \ \& \ k \leq m; \\ d_{i-n,k-m}, & \text{if } i > n \ \& \ k > m \end{cases} \begin{cases} a'_{k,j}, & \text{if } k \leq m \ \& \ j \leq \ell; \\ b'_{k,j-\ell}, & \text{if } k \leq m \ \& \ j > \ell; \\ c'_{k-m,j}, & \text{if } k > m \ \& \ j \leq \ell; \\ d'_{k-m,j-\ell}, & \text{if } k > m \ \& \ j > \ell \end{cases} \right)_{1\leq i\leq n+n',\ 1\leq j\leq \ell+\ell'} . \tag{1215}$$

On the other hand, we have $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$ and $A' = \left(a'_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq \ell}$. Hence, the definition of the product of two matrices shows that

$$AA' = \left(\sum_{k=1}^{m} a_{i,k} a'_{k,j} \right)_{1\leq i\leq n,\ 1\leq j\leq \ell} . \tag{1216}$$

Also, we have $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m'}$ and $C' = \left(c'_{i,j}\right)_{1\leq i\leq m',\ 1\leq j\leq \ell}$. Hence, the definition of the product of two matrices shows that

$$BC' = \left(\sum_{k=1}^{m'} b_{i,k} c'_{k,j} \right)_{1\leq i\leq n,\ 1\leq j\leq \ell} . \tag{1217}$$

Adding the equalities (1216) and (1217), we obtain

$$\begin{aligned} AA' + BC' &= \left(\sum_{k=1}^{m} a_{i,k} a'_{k,j} \right)_{1 \le i \le n,\ 1 \le j \le \ell} + \left(\sum_{k=1}^{m'} b_{i,k} c'_{k,j} \right)_{1 \le i \le n,\ 1 \le j \le \ell} \\ &= \left(\sum_{k=1}^{m} a_{i,k} a'_{k,j} + \underbrace{\sum_{k=1}^{m'} b_{i,k} c'_{k,j}}_{\substack{= \sum_{k=m+1}^{m+m'} b_{i,k-m} c'_{k-m,j} \\ \text{(here, we have substituted } k-m \text{ for } k \text{ in the sum)}}} \right)_{1 \le i \le n,\ 1 \le j \le \ell} \\ &= \left(\sum_{k=1}^{m} a_{i,k} a'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m} c'_{k-m,j} \right)_{1 \le i \le n,\ 1 \le j \le \ell} . \end{aligned} \tag{1218}$$

Similarly,

$$AB' + BD' = \left(\sum_{k=1}^{m} a_{i,k} b'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m} d'_{k-m,j} \right)_{1 \le i \le n,\ 1 \le j \le \ell'} ; \tag{1219}$$

$$CA' + DC' = \left(\sum_{k=1}^{m} c_{i,k} a'_{k,j} + \sum_{k=m+1}^{m+m'} d_{i,k-m} c'_{k-m,j} \right)_{1 \le i \le n',\ 1 \le j \le \ell} ; \tag{1220}$$

$$CB' + DD' = \left(\sum_{k=1}^{m} c_{i,k} b'_{k,j} + \sum_{k=m+1}^{m+m'} d_{i,k-m} d'_{k-m,j} \right)_{1 \le i \le n',\ 1 \le j \le \ell'} . \tag{1221}$$

Now, we have the four equalities (1218), (1219), (1220) and (1221). Hence, the definition of the block matrix $\begin{pmatrix} AA' + BC' & AB' + BD' \\ CA' + DC' & CB' + DD' \end{pmatrix}$ (or, more precisely, the equality (435), applied to n, n', ℓ, ℓ', $AA' + BC'$, $AB' + BD'$, $CA' + DC'$, $CB' + DD'$, $\sum_{k=1}^{m} a_{i,k} a'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m} c'_{k-m,j}$, $\sum_{k=1}^{m} a_{i,k} b'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m} d'_{k-m,j}$, $\sum_{k=1}^{m} c_{i,k} a'_{k,j} + \sum_{k=m+1}^{m+m'} d_{i,k-m} c'_{k-m,j}$ and $\sum_{k=1}^{m} c_{i,k} b'_{k,j} + \sum_{k=m+1}^{m+m'} d_{i,k-m} d'_{k-m,j}$ instead of n, n', m,

m', A, B, C, D, $a_{i,j}$, $b_{i,j}$, $c_{i,j}$ and $d_{i,j}$) shows that

$$\begin{pmatrix} AA'+BC' & AB'+BD' \\ CA'+DC' & CB'+DD' \end{pmatrix}$$
$$= \left(\begin{cases} \sum_{k=1}^{m} a_{i,k}a'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m}c'_{k-m,j}, & \text{if } i \leq n \ \& \ j \leq \ell; \\ \sum_{k=1}^{m} a_{i,k}b'_{k,j-\ell} + \sum_{k=m+1}^{m+m'} b_{i,k-m}d'_{k-m,j-\ell}, & \text{if } i \leq n \ \& \ j > \ell; \\ \sum_{k=1}^{m} c_{i-n,k}a'_{k,j} + \sum_{k=m+1}^{m+m'} d_{i-n,k-m}c'_{k-m,j}, & \text{if } i > n \ \& \ j \leq \ell; \\ \sum_{k=1}^{m} c_{i-n,k}b'_{k,j-\ell} + \sum_{k=m+1}^{m+m'} d_{i-n,k-m}d'_{k-m,j-\ell}, & \text{if } i > n \ \& \ j > \ell \end{cases} \right)_{1\leq i\leq n+n',\ 1\leq j\leq \ell+\ell'} . \tag{1222}$$

But our goal is to prove that $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} AA'+BC' & AB'+BD' \\ CA'+DC' & CB'+DD' \end{pmatrix}$. In other words, we want to prove that the left hand sides of the equalities (1215) and (1222) are equal. For this, it clearly suffices to show that the right hand sides of these equalities are equal. In other words, it suffices to show that every $(i,j) \in \{1,2,\ldots,n+n'\} \times \{1,2,\ldots,\ell+\ell'\}$ satisfies

$$\sum_{k=1}^{m+m'} \begin{cases} a_{i,k}, & \text{if } i \leq n \ \& \ k \leq m; \\ b_{i,k-m}, & \text{if } i \leq n \ \& \ k > m; \\ c_{i-n,k}, & \text{if } i > n \ \& \ k \leq m; \\ d_{i-n,k-m}, & \text{if } i > n \ \& \ k > m \end{cases} \begin{cases} a'_{k,j}, & \text{if } k \leq m \ \& \ j \leq \ell; \\ b'_{k,j-\ell}, & \text{if } k \leq m \ \& \ j > \ell; \\ c'_{k-m,j}, & \text{if } k > m \ \& \ j \leq \ell; \\ d'_{k-m,j-\ell}, & \text{if } k > m \ \& \ j > \ell \end{cases}$$
$$= \begin{cases} \sum_{k=1}^{m} a_{i,k}a'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m}c'_{k-m,j}, & \text{if } i \leq n \ \& \ j \leq \ell; \\ \sum_{k=1}^{m} a_{i,k}b'_{k,j-\ell} + \sum_{k=m+1}^{m+m'} b_{i,k-m}d'_{k-m,j-\ell}, & \text{if } i \leq n \ \& \ j > \ell; \\ \sum_{k=1}^{m} c_{i-n,k}a'_{k,j} + \sum_{k=m+1}^{m+m'} d_{i-n,k-m}c'_{k-m,j}, & \text{if } i > n \ \& \ j \leq \ell; \\ \sum_{k=1}^{m} c_{i-n,k}b'_{k,j-\ell} + \sum_{k=m+1}^{m+m'} d_{i-n,k-m}d'_{k-m,j-\ell}, & \text{if } i > n \ \& \ j > \ell \end{cases} . \tag{1223}$$

[*Proof of (1223):* Let $(i,j) \in \{1,2,\ldots,n+n'\} \times \{1,2,\ldots,\ell+\ell'\}$. Thus, $i \in \{1,2,\ldots,n+n'\}$ and $j \in \{1,2,\ldots,\ell+\ell'\}$. We must be in one of the following four cases:

Case 1: We have $i \leq n$ and $j \leq \ell$.

Case 2: We have $i \leq n$ and $j > \ell$.

Case 3: We have $i > n$ and $j \leq \ell$.

Case 4: We have $i > n$ and $j > \ell$.

All four cases are completely analogous; we thus will only show how to deal with Case 1. In this case, we have $i \leq n$ and $j \leq \ell$. Now, comparing

$$\sum_{k=1}^{m+m'} \begin{cases} a_{i,k}, & \text{if } i \leq n \ \& \ k \leq m; \\ b_{i,k-m}, & \text{if } i \leq n \ \& \ k > m; \\ c_{i-n,k}, & \text{if } i > n \ \& \ k \leq m; \\ d_{i-n,k-m}, & \text{if } i > n \ \& \ k > m \end{cases} \begin{cases} a'_{k,j}, & \text{if } k \leq m \ \& \ j \leq \ell; \\ b'_{k,j-\ell}, & \text{if } k \leq m \ \& \ j > \ell; \\ c'_{k-m,j}, & \text{if } k > m \ \& \ j \leq \ell; \\ d'_{k-m,j-\ell}, & \text{if } k > m \ \& \ j > \ell \end{cases}$$

$$= \sum_{k=1}^{m} \underbrace{\begin{cases} a_{i,k}, & \text{if } i \leq n \ \& \ k \leq m; \\ b_{i,k-m}, & \text{if } i \leq n \ \& \ k > m; \\ c_{i-n,k}, & \text{if } i > n \ \& \ k \leq m; \\ d_{i-n,k-m}, & \text{if } i > n \ \& \ k > m \end{cases}}_{\substack{=a_{i,k} \\ (\text{since } i \leq n \text{ and } k \leq m)}} \underbrace{\begin{cases} a'_{k,j}, & \text{if } k \leq m \ \& \ j \leq \ell; \\ b'_{k,j-\ell}, & \text{if } k \leq m \ \& \ j > \ell; \\ c'_{k-m,j}, & \text{if } k > m \ \& \ j \leq \ell; \\ d'_{k-m,j-\ell}, & \text{if } k > m \ \& \ j > \ell \end{cases}}_{\substack{=a'_{k,j} \\ (\text{since } k \leq m \text{ and } j \leq \ell)}}$$

$$+ \sum_{k=m+1}^{m+m'} \underbrace{\begin{cases} a_{i,k}, & \text{if } i \leq n \ \& \ k \leq m; \\ b_{i,k-m}, & \text{if } i \leq n \ \& \ k > m; \\ c_{i-n,k}, & \text{if } i > n \ \& \ k \leq m; \\ d_{i-n,k-m}, & \text{if } i > n \ \& \ k > m \end{cases}}_{\substack{=b_{i,k-m} \\ (\text{since } i \leq n \text{ and } k > m)}} \underbrace{\begin{cases} a'_{k,j}, & \text{if } k \leq m \ \& \ j \leq \ell; \\ b'_{k,j-\ell}, & \text{if } k \leq m \ \& \ j > \ell; \\ c'_{k-m,j}, & \text{if } k > m \ \& \ j \leq \ell; \\ d'_{k-m,j-\ell}, & \text{if } k > m \ \& \ j > \ell \end{cases}}_{\substack{=c'_{k-m,j} \\ (\text{since } k > m \text{ and } j \leq \ell)}}$$

$$\left(\text{since } 0 \leq m \leq m + m'\right)$$

$$= \sum_{k=1}^{m} a_{i,k} a'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m} c'_{k-m,j}$$

with

$$\begin{cases} \sum\limits_{k=1}^{m} a_{i,k} a'_{k,j} + \sum\limits_{k=m+1}^{m+m'} b_{i,k-m} c'_{k-m,j}, & \text{if } i \leq n \text{ \& } j \leq \ell; \\ \sum\limits_{k=1}^{m} a_{i,k} b'_{k,j-\ell} + \sum\limits_{k=m+1}^{m+m'} b_{i,k-m} d'_{k-m,j-\ell}, & \text{if } i \leq n \text{ \& } j > \ell; \\ \sum\limits_{k=1}^{m} c_{i-n,k} a'_{k,j} + \sum\limits_{k=m+1}^{m+m'} d_{i-n,k-m} c'_{k-m,j}, & \text{if } i > n \text{ \& } j \leq \ell; \\ \sum\limits_{k=1}^{m} c_{i-n,k} b'_{k,j-\ell} + \sum\limits_{k=m+1}^{m+m'} d_{i-n,k-m} d'_{k-m,j-\ell}, & \text{if } i > n \text{ \& } j > \ell \end{cases}$$

$$= \sum_{k=1}^{m} a_{i,k} a'_{k,j} + \sum_{k=m+1}^{m+m'} b_{i,k-m} c'_{k-m,j} \qquad (\text{since } i \leq n \text{ and } j \leq \ell),$$

we obtain precisely (1223). Thus, (1223) is proven in Case 1. As I said, the other three cases are similar, and so (1223) is proven.]

From (1223), we see that the right hand sides of the equalities (1215) and (1222) are equal. Hence, so are their left hand sides. In other words, $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} AA' + BC' & AB' + BD' \\ CA' + DC' & CB' + DD' \end{pmatrix}$. This solves Exercise 6.28. □

7.97. Solution to Exercise 6.29

Solution to Exercise 6.29. We shall prove Exercise 6.29 by induction over m:

Induction base: Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let B be an $n \times 0$-matrix[544]. Let D be a 0×0-matrix[545]. Then, all three matrices B, $0_{0\times n}$ and D are empty (in the sense that each of them has either 0 rows or 0 columns or both), and thus we have $\begin{pmatrix} A & B \\ 0_{0\times n} & D \end{pmatrix} = A$. Hence, $\det \begin{pmatrix} A & B \\ 0_{0\times n} & D \end{pmatrix} = \det A$. Combined with $\det D = 1$ (since D is a 0×0-matrix), this yields $\det \begin{pmatrix} A & B \\ 0_{0\times n} & D \end{pmatrix} = \det A \cdot \det D$.

Now, let us forget that we fixed n, A, B and D. We thus have proven that every $n \in \mathbb{N}$, every $n \times n$-matrix A, every $n \times 0$-matrix B and every 0×0-matrix D satisfy $\det \begin{pmatrix} A & B \\ 0_{0\times n} & D \end{pmatrix} = \det A \cdot \det D$. In other words, Exercise 6.29 holds for $m = 0$. This completes the induction base.

Induction step: Let $M \in \mathbb{N}$ be positive. Assume that Exercise 6.29 holds for $m = M - 1$. We need to prove that Exercise 6.29 holds for $m = M$.

Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let B be an $n \times M$-matrix. Let D be an $M \times M$-matrix.

[544]Of course, there is only one such $n \times 0$-matrix (namely, the empty matrix).
[545]Of course, there is only one such 0×0-matrix (namely, the empty matrix).

Write the $M \times M$-matrix D in the form $D = \left(d_{i,j}\right)_{1 \leq i \leq M,\ 1 \leq j \leq M}$. Hence, Theorem 6.82 **(a)** (applied to M, D, $d_{i,j}$ and M instead of n, A, $a_{i,j}$ and p) shows that

$$\det D = \sum_{q=1}^{M} (-1)^{M+q} d_{M,q} \det \left(D_{\sim M, \sim q}\right) \tag{1224}$$

(since $M \in \{1, 2, \ldots, M\}$ (because $M > 0$)).

Write the $(n+M) \times (n+M)$-matrix $\begin{pmatrix} A & B \\ 0_{M \times n} & D \end{pmatrix}$ in the form $\begin{pmatrix} A & B \\ 0_{M \times n} & D \end{pmatrix} = \left(u_{i,j}\right)_{1 \leq u \leq n+M,\ 1 \leq v \leq n+M}$. Thus,

$$u_{n+M,q} = 0 \qquad \text{for every } q \in \{1, 2, \ldots, n\}, \tag{1225}$$

and

$$u_{n+M,n+q} = d_{M,q} \qquad \text{for every } q \in \{1, 2, \ldots, M\}. \tag{1226}$$

Furthermore, for every $q \in \{1, 2, \ldots, M\}$, we have

$$\begin{pmatrix} A & B \\ 0_{M \times n} & D \end{pmatrix}_{\sim(n+M), \sim(n+q)} = \begin{pmatrix} A & B'_q \\ 0_{(M-1) \times n} & D_{\sim M, \sim q} \end{pmatrix}, \tag{1227}$$

where B'_q is the result of crossing out the q-th column in the matrix B. (Draw the matrices and cross out the appropriate rows and columns to see why this is true.)

But we have assumed that Exercise 6.29 holds for $m = M - 1$. Hence, for every $q \in \{1, 2, \ldots, M\}$, we can apply Exercise 6.29 to $M - 1$, B'_q and $D_{\sim M, \sim q}$ instead of m, B and D. As a result, for every $q \in \{1, 2, \ldots, M\}$, we obtain

$$\det \begin{pmatrix} A & B'_q \\ 0_{(M-1) \times n} & D_{\sim M, \sim q} \end{pmatrix} = \det A \cdot \det \left(D_{\sim M, \sim q}\right).$$

Now, taking determinants in (1227), we obtain

$$\begin{aligned} \det \left(\begin{pmatrix} A & B \\ 0_{M \times n} & D \end{pmatrix}_{\sim(n+M), \sim(n+q)} \right) &= \det \begin{pmatrix} A & B'_q \\ 0_{(M-1) \times n} & D_{\sim M, \sim q} \end{pmatrix} \\ &= \det A \cdot \det \left(D_{\sim M, \sim q}\right). \end{aligned} \tag{1228}$$

But $n + M > 0$ and thus $n + M \in \{1, 2, \ldots, n + M\}$. Thus, Theorem 6.82 **(a)** (applied to $n + M$, $\begin{pmatrix} A & B \\ 0_{M \times n} & D \end{pmatrix}$, $u_{i,j}$ and $n + M$ instead of n, A, $a_{i,j}$ and p) shows that

$$\begin{aligned}
&\det \begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix} \\
&= \sum_{q=1}^{n+M} (-1)^{(n+M)+q} u_{n+M,q} \det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim q} \right) \\
&= \sum_{q=1}^{n} (-1)^{(n+M)+q} \underbrace{u_{n+M,q}}_{\substack{=0 \\ \text{(by (1225))}}} \det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim q} \right) \\
&\qquad + \sum_{q=n+1}^{n+M} (-1)^{(n+M)+q} u_{n+M,q} \det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim q} \right) \\
&\qquad (\text{since } 0 \leq n \leq n+M) \\
&= \underbrace{\sum_{q=1}^{n} (-1)^{(n+M)+q} 0 \det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim q} \right)}_{=0} \\
&\qquad + \sum_{q=n+1}^{n+M} (-1)^{(n+M)+q} u_{n+M,q} \det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim q} \right) \\
&= \sum_{q=n+1}^{n+M} (-1)^{(n+M)+q} u_{n+M,q} \det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim q} \right) \\
&= \sum_{q=1}^{M} \underbrace{(-1)^{(n+M)+(n+q)}}_{\substack{=(-1)^{M+q} \\ \text{(since } (n+M)+(n+q) \\ =2n+M+q\equiv M+q \operatorname{mod} 2)}} \underbrace{u_{n+M,n+q}}_{\substack{=d_{M,q} \\ \text{(by (1226))}}} \underbrace{\det \left(\begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix}_{\sim(n+M),\sim(n+q)} \right)}_{\substack{=\det A\cdot\det\left(D_{\sim M,\sim q}\right) \\ \text{(by (1228))}}} \\
&\qquad (\text{here, we have substituted } n+q \text{ for } q \text{ in the sum}) \\
&= \sum_{q=1}^{M} (-1)^{M+q} d_{M,q} \det A \cdot \det\left(D_{\sim M,\sim q}\right) = \det A \cdot \underbrace{\sum_{q=1}^{M} (-1)^{M+q} d_{M,q} \det\left(D_{\sim M,\sim q}\right)}_{\substack{=\det D \\ \text{(by (1224))}}} \\
&= \det A \cdot \det D.
\end{aligned}$$

Now, let us forget that we fixed n, A, B and D. We thus have shown that for every $n \in \mathbb{N}$, for every $n\times n$-matrix A, for every $n\times M$-matrix B, and for every $M\times M$-matrix D, we have $\det \begin{pmatrix} A & B \\ 0_{M\times n} & D \end{pmatrix} = \det A \cdot \det D$. In other words, Exercise 6.29 holds for $m = M$. This completes the induction step. Hence, Exercise 6.29 is solved by induction. □

7.98. Solution to Exercise 6.30

There are two ways to solve Exercise 6.30: One way is to essentially repeat our above solution to Exercise 6.29 with some straightforward modifications (for example, we must use Theorem 6.82 **(b)** instead of Theorem 6.82 **(a)**). Another way is to derive Exercise 6.30 from Exercise 6.29 using transpose matrices. Let me show the latter way. We begin with a simple lemma:

Lemma 7.173. Let n, n', m and m' be four nonnegative integers. Let $A \in \mathbb{K}^{n\times m}$, $B \in \mathbb{K}^{n\times m'}$, $C \in \mathbb{K}^{n'\times m}$ and $D \in \mathbb{K}^{n'\times m'}$. Then,

$$\left(\begin{array}{cc} A & B \\ C & D \end{array}\right)^T = \left(\begin{array}{cc} A^T & C^T \\ B^T & D^T \end{array}\right).$$

Proof of Lemma 7.173. Lemma 7.173 results in a straightforward way by recalling the definitions of $\left(\begin{array}{cc} A & B \\ C & D \end{array}\right)^T$ and $\left(\begin{array}{cc} A^T & C^T \\ B^T & D^T \end{array}\right)$ and comparing. □

Now, we can comfortably solve Exercise 6.30:

Solution to Exercise 6.30. Exercise 6.4 (applied to $n+m$ and $\left(\begin{array}{cc} A & 0_{n\times m} \\ C & D \end{array}\right)$ instead of n and A) shows that

$$\det\left(\left(\begin{array}{cc} A & 0_{n\times m} \\ C & D \end{array}\right)^T\right) = \det\left(\begin{array}{cc} A & 0_{n\times m} \\ C & D \end{array}\right).$$

Hence,

$$\begin{aligned}
&\det\begin{pmatrix} A & 0_{n\times m} \\ C & D \end{pmatrix} \\
&= \det\left(\underbrace{\begin{pmatrix} A & 0_{n\times m} \\ C & D \end{pmatrix}^T}_{\substack{=\begin{pmatrix} A^T & C^T \\ (0_{n\times m})^T & D^T \end{pmatrix} \\ \text{(by Lemma 7.173 (applied to } n,\ m,\ n,\ m \\ \text{and } 0_{n\times m} \text{ instead of } n,\ n',\ m,\ m' \text{ and } B))}} \right) = \det \underbrace{\begin{pmatrix} A^T & C^T \\ (0_{n\times m})^T & D^T \end{pmatrix}}_{\substack{=\begin{pmatrix} A^T & C^T \\ 0_{m\times n} & D^T \end{pmatrix} \\ \text{(since } (0_{n\times m})^T = 0_{m\times n})}} \\
&= \det\begin{pmatrix} A^T & C^T \\ 0_{m\times n} & D^T \end{pmatrix} = \underbrace{\det\left(A^T\right)}_{\substack{=\det A \\ \text{(by Exercise 6.4)}}} \cdot \underbrace{\det\left(D^T\right)}_{\substack{=\det D \\ \text{(by Exercise 6.4} \\ \text{(applied to } m \text{ and } D \text{ instead} \\ \text{of } n \text{ and } A))}} \\
&\qquad \left(\text{by Exercise 6.29 (applied to } A^T,\ C^T \text{ and } D^T \text{ instead of } A,\ B \text{ and } D)\right) \\
&= \det A \cdot \det D.
\end{aligned}$$

This solves Exercise 6.30. □

7.99. Second solution to Exercise 6.6

Second solution to Exercise 6.6 (sketched). **(b)** We have

$$\begin{aligned}
&\det\begin{pmatrix} a & b & c & d & e \\ f & 0 & 0 & 0 & g \\ h & 0 & 0 & 0 & i \\ j & 0 & 0 & 0 & k \\ \ell & m & n & o & p \end{pmatrix} \\
&= -\det\begin{pmatrix} a & b & c & d & e \\ \ell & m & n & o & p \\ h & 0 & 0 & 0 & i \\ j & 0 & 0 & 0 & k \\ f & 0 & 0 & 0 & g \end{pmatrix} \\
&\qquad \left(\begin{array}{c}\text{by Exercise 6.7 \textbf{(a)}, because we have just} \\ \text{switched the 2-nd and the 5-th rows of the matrix}\end{array}\right) \\
&= \det\begin{pmatrix} d & b & c & a & e \\ o & m & n & \ell & p \\ 0 & 0 & 0 & h & i \\ 0 & 0 & 0 & j & k \\ 0 & 0 & 0 & f & g \end{pmatrix} \\
&\qquad \left(\begin{array}{c}\text{by Exercise 6.7 \textbf{(b)}, because we have just} \\ \text{switched the 1-st and the 4-th columns of the matrix}\end{array}\right) \\
&= \underbrace{\det\begin{pmatrix} d & b & c \\ o & m & n \\ 0 & 0 & 0 \end{pmatrix}}_{\substack{=0 \\ \text{(by Exercise 6.7 \textbf{(c)})}}} \cdot \det\begin{pmatrix} j & k \\ f & g \end{pmatrix} \\
&\qquad \left(\begin{array}{c}\text{by Exercise 6.29, applied to } \begin{pmatrix} d & b & c \\ o & m & n \\ 0 & 0 & 0 \end{pmatrix}, \\ \begin{pmatrix} a & e \\ \ell & p \\ h & i \end{pmatrix}, \begin{pmatrix} j & k \\ f & g \end{pmatrix}, 2 \text{ and } 3 \text{ instead of } A, B, D, m \text{ and } n\end{array}\right) \\
&= 0.
\end{aligned}$$

This solves Exercise 6.6 **(b)**.

(a) We have

$$\det \begin{pmatrix} a & b & c & d \\ \ell & 0 & 0 & e \\ k & 0 & 0 & f \\ j & i & h & g \end{pmatrix}$$
$$= -\det \begin{pmatrix} a & b & c & d \\ j & i & h & g \\ k & 0 & 0 & f \\ \ell & 0 & 0 & e \end{pmatrix}$$
$$\left(\begin{array}{c} \text{by Exercise 6.7 } \textbf{(a)}\text{, because we have just} \\ \text{switched the 2-nd and the 4-th rows of the matrix} \end{array} \right)$$
$$= \det \begin{pmatrix} c & b & a & d \\ h & i & j & g \\ 0 & 0 & k & f \\ 0 & 0 & \ell & e \end{pmatrix}$$
$$\left(\begin{array}{c} \text{by Exercise 6.7 } \textbf{(b)}\text{, because we have just} \\ \text{switched the 1-st and the 3-th columns of the matrix} \end{array} \right)$$
$$= \underbrace{\det \begin{pmatrix} c & b \\ h & i \end{pmatrix}}_{=ci-bh} \cdot \underbrace{\det \begin{pmatrix} k & f \\ \ell & e \end{pmatrix}}_{=ek-\ell f}$$
$$\left(\begin{array}{c} \text{by Exercise 6.29, applied to } \begin{pmatrix} c & b \\ h & i \end{pmatrix}, \\ \begin{pmatrix} a & d \\ j & g \end{pmatrix}, \begin{pmatrix} k & f \\ \ell & e \end{pmatrix}, \text{2 and 2 instead of } A, B, D, m \text{ and } n \end{array} \right)$$
$$= (ci - bh)(ek - \ell f) = (bh - ci)(\ell f - ek).$$

This solves Exercise 6.6 **(a)**. (Notice that we have obtained the result in its factored form!) □

7.100. Solution to Exercise 6.31

Before we solve Exercise 6.31, let us introduce some notation that enables us to clearly speak about permutations of rows and columns in a matrix:

Definition 7.174. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m} \in \mathbb{K}^{n \times m}$. Let $\gamma \in S_n$ and $\delta \in S_m$. Then, $A_{[\gamma,\delta]}$ denotes the matrix $\left(a_{\gamma(i),\delta(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m} \in \mathbb{K}^{n \times m}$.

We shall use this definition throughout Section 7.100.

Example 7.175. For this example, let $n = 3$ and $m = 4$, and let $A = \begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{pmatrix} \in \mathbb{K}^{3\times 4}$ be a 3×4-matrix. Let $\gamma \in S_3$ be the permutation that sends $1,2,3$ to $2,3,1$, respectively. Let $\delta \in S_4$ be the permutation that sends $1,2,3,4$ to $2,1,4,3$, respectively. Then,

$$A_{[\gamma,\delta]} = \begin{pmatrix} b' & a' & d' & c' \\ b'' & a'' & d'' & c'' \\ b & a & d & c \end{pmatrix}.$$

Remark 7.176. Let n, m, A, γ and δ be as in Definition 7.174. Then, it is easy to see that $A_{[\gamma,\delta]} = \operatorname{sub}^{\delta(1),\delta(2),\ldots,\delta(m)}_{\gamma(1),\gamma(2),\ldots,\gamma(n)} A$ (where we are using the notation introduced in Definition 6.78). Visually speaking, $A_{[\gamma,\delta]}$ is the matrix obtained from A by permuting the rows (using the permutation γ) and permuting the columns (using the permutation δ).

Remark 7.177. Let $n \in \mathbb{N}$, and let $B \in \mathbb{K}^{n\times n}$. Let $\kappa \in S_n$. Then, Definition 7.174 gives rise to an $n \times n$-matrix $B_{[\kappa,\mathrm{id}]}$ (where id denotes the identity permutation $\mathrm{id}_{\{1,2,\ldots,n\}} \in S_n$). This matrix $B_{[\kappa,\mathrm{id}]}$ is precisely the matrix B_κ from Lemma 6.17.

The following lemma will be crucial for us:

Lemma 7.178. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times n}$. Let $\gamma \in S_n$ and $\delta \in S_n$. Then,

$$\det\left(A_{[\gamma,\delta]}\right) = (-1)^\gamma (-1)^\delta \det A.$$

Lemma 7.178 simply says that if we permute the rows and permute the columns of a square matrix, then the determinant of this matrix gets multiplied by the product of the signs of the two permutations.

Proof of Lemma 7.178. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$.

Write the $n \times n$-matrix A in the form $A = (a_{i,j})_{1\le i\le n,\ 1\le j\le n}$. Then, $A^T = (a_{j,i})_{1\le i\le n,\ 1\le j\le n}$ (by the definition of A^T).

Define a new matrix C by $C = \left(a_{j,\delta(i)}\right)_{1\le i\le n,\ 1\le j\le n}$. Then, Lemma 6.17 **(a)** (applied to δ, A^T, $a_{j,i}$ and C instead of κ, B, $b_{i,j}$ and B_κ) yields $\det C = (-1)^\delta \cdot \det\left(A^T\right)$ (because $A^T = (a_{j,i})_{1\le i\le n,\ 1\le j\le n}$ and $C = \left(a_{j,\delta(i)}\right)_{1\le i\le n,\ 1\le j\le n}$). Thus,

$$\det C = (-1)^\delta \cdot \underbrace{\det\left(A^T\right)}_{\substack{=\det A \\ \text{(by Exercise 6.4)}}} = (-1)^\delta \cdot \det A.$$

Clearly, C is an $n \times n$-matrix. Hence, Exercise 6.4 (applied to C instead of A) yields

$$\det\left(C^T\right) = \det C = (-1)^\delta \cdot \det A.$$

But we have $C = \left(a_{j,\delta(i)}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Hence, $C^T = \left(a_{i,\delta(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of C^T). We have $C^T = \left(a_{i,\delta(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $A_{[\gamma,\delta]} = \left(a_{\gamma(i),\delta(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of $A_{[\gamma,\delta]}$). Thus, Lemma 6.17 **(a)** (applied to γ, C^T, $a_{i,\delta(j)}$ and $A_{[\gamma,\delta]}$ instead of κ, B, $b_{i,j}$ and B_κ) yields

$$\det\left(A_{[\gamma,\delta]}\right) = (-1)^\gamma \cdot \underbrace{\det\left(C^T\right)}_{=(-1)^\delta\cdot\det A} = (-1)^\gamma \cdot (-1)^\delta \cdot \det A = (-1)^\gamma (-1)^\delta \det A.$$

This proves Lemma 7.178. □

Corollary 7.179. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times n}$. Let $\gamma \in S_n$. Then, $\det\left(A_{[\gamma,\gamma]}\right) = \det A$.

Proof of Corollary 7.179. We have $(-1)^\gamma = (-1)^{\ell(\gamma)}$ (by the definition of $(-1)^\gamma$). Hence,

$$\left(\underbrace{(-1)^\gamma}_{=(-1)^{\ell(\gamma)}}\right)^2 = \left((-1)^{\ell(\gamma)}\right)^2 = (-1)^{\ell(\gamma)\cdot 2} = 1$$

(since the integer $\ell(\gamma)\cdot 2$ is even). But Lemma 7.178 (applied to $\delta = \gamma$) yields

$$\det\left(A_{[\gamma,\gamma]}\right) = \underbrace{(-1)^\gamma (-1)^\gamma}_{=\left((-1)^\gamma\right)^2=1} \det A = \det A.$$

This proves Corollary 7.179. □

Solution to Exercise 6.31. **(a)** Let $a, a', a'', b, b', b'', c, c', c'', d, d', d'', e$ be elements of $\mathbb{K}$. Let A be the 7×7-matrix

$$\begin{pmatrix} a & 0 & 0 & 0 & 0 & 0 & b \\ 0 & a' & 0 & 0 & 0 & b' & 0 \\ 0 & 0 & a'' & 0 & b'' & 0 & 0 \\ 0 & 0 & 0 & e & 0 & 0 & 0 \\ 0 & 0 & c'' & 0 & d'' & 0 & 0 \\ 0 & c' & 0 & 0 & 0 & d' & 0 \\ c & 0 & 0 & 0 & 0 & 0 & d \end{pmatrix}.$$

Thus, the exercise demands that we find $\det A$.

Let $\gamma \in S_7$ be the permutation of the set $\{1,2,3,4,5,6,7\}$ which sends $1,2,3,4,5,6,7$ to $1,7,3,5,4,6,2$, respectively. (This permutation γ swaps 2 with 7 and swaps 4 with 5; it leaves $1,3,6$ unchanged.) Then, straightforward computation (using the definitions of A and of $A_{[\gamma,\gamma]}$) shows that

$$A_{[\gamma,\gamma]} = \begin{pmatrix} a & b & 0 & 0 & 0 & 0 & 0 \\ c & d & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a'' & b'' & 0 & 0 & 0 \\ 0 & 0 & c'' & d'' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & d' & c' \\ 0 & 0 & 0 & 0 & 0 & b' & a' \end{pmatrix}. \tag{1229}$$

But Corollary 7.179 (applied to $n = 7$) yields $\det\left(A_{[\gamma,\gamma]}\right) = \det A$. Hence,

$$\begin{aligned} \det A = \det\left(A_{[\gamma,\gamma]}\right) &= \det \begin{pmatrix} a & b & 0 & 0 & 0 & 0 & 0 \\ c & d & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a'' & b'' & 0 & 0 & 0 \\ 0 & 0 & c'' & d'' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & d' & c' \\ 0 & 0 & 0 & 0 & 0 & b' & a' \end{pmatrix} \qquad (\text{by } (1229)) \\ &= \det \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a'' & b'' \\ 0 & 0 & c'' & d'' \end{pmatrix} \cdot \det \begin{pmatrix} e & 0 & 0 \\ 0 & d' & c' \\ 0 & b' & a' \end{pmatrix} \end{aligned} \tag{1230}$$

(by Exercise 6.29, applied to $\begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a'' & b'' \\ 0 & 0 & c'' & d'' \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} e & 0 & 0 \\ 0 & d' & c' \\ 0 & b' & a' \end{pmatrix}$, 3 and 4 instead of A, B, D, m and n).

But Exercise 6.29 (applied to $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} a'' & b'' \\ c'' & d'' \end{pmatrix}$, 2 and 2 instead of A, B, D, m and n) yields

$$\det \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a'' & b'' \\ 0 & 0 & c'' & d'' \end{pmatrix} = \underbrace{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{=ad-bc} \cdot \underbrace{\det \begin{pmatrix} a'' & b'' \\ c'' & d'' \end{pmatrix}}_{=a''d''-b''c''} = (ad-bc) \cdot \left(a''d'' - b''c''\right).$$

Also, Exercise 6.29 (applied to $\begin{pmatrix} e \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \end{pmatrix}$, $\begin{pmatrix} d' & c' \\ b' & a' \end{pmatrix}$, 2 and 1 instead of A, B,

D, m and n) yields

$$\det\begin{pmatrix} e & 0 & 0 \\ 0 & d' & c' \\ 0 & b' & a' \end{pmatrix} = \underbrace{\det\begin{pmatrix} e \end{pmatrix}}_{=e} \cdot \underbrace{\det\begin{pmatrix} d' & c' \\ b' & a' \end{pmatrix}}_{\substack{=d'a'-c'b' \\ =a'd'-b'c'}} = e \cdot \left(a'd' - b'c'\right).$$

Hence, (1230) becomes

$$\det A = \underbrace{\det\begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a'' & b'' \\ 0 & 0 & c'' & d'' \end{pmatrix}}_{=(ad-bc)\cdot(a''d''-b''c'')} \cdot \underbrace{\det\begin{pmatrix} e & 0 & 0 \\ 0 & d' & c' \\ 0 & b' & a' \end{pmatrix}}_{=e\cdot(a'd'-b'c')}$$
$$= (ad-bc)\cdot\left(a''d''-b''c''\right)\cdot e\cdot\left(a'd'-b'c'\right)$$
$$= e\cdot(ad-bc)\cdot\left(a''d''-b''c''\right)\cdot\left(a'd'-b'c'\right).$$

This solves Exercise 6.31 **(a)**.

(b) Let $a, b, c, d, e, f, g, h, k, \ell, m, n$ be elements of $\mathbb{K}$. Let A be the 6×6-matrix

$$\begin{pmatrix} a & 0 & 0 & \ell & 0 & 0 \\ 0 & b & 0 & 0 & m & 0 \\ 0 & 0 & c & 0 & 0 & n \\ g & 0 & 0 & d & 0 & 0 \\ 0 & h & 0 & 0 & e & 0 \\ 0 & 0 & k & 0 & 0 & f \end{pmatrix}.$$

Thus, the exercise demands that we find $\det A$.

Let $\gamma \in S_6$ be the permutation of the set $\{1,2,3,4,5,6\}$ which sends $1,2,3,4,5,6$ to $1,4,2,5,3,6$, respectively. Then, straightforward computation (using the definitions of A and of $A_{[\gamma,\gamma]}$) shows that

$$A_{[\gamma,\gamma]} = \begin{pmatrix} a & \ell & 0 & 0 & 0 & 0 \\ g & d & 0 & 0 & 0 & 0 \\ 0 & 0 & b & m & 0 & 0 \\ 0 & 0 & h & e & 0 & 0 \\ 0 & 0 & 0 & 0 & c & n \\ 0 & 0 & 0 & 0 & k & f \end{pmatrix}. \qquad (1231)$$

But Corollary 7.179 (applied to $n = 6$) yields $\det\left(A_{[\gamma,\gamma]}\right) = \det A$. Hence,

$$\det A = \det\left(A_{[\gamma,\gamma]}\right) = \det\begin{pmatrix} a & \ell & 0 & 0 & 0 & 0 \\ g & d & 0 & 0 & 0 & 0 \\ 0 & 0 & b & m & 0 & 0 \\ 0 & 0 & h & e & 0 & 0 \\ 0 & 0 & 0 & 0 & c & n \\ 0 & 0 & 0 & 0 & k & f \end{pmatrix} \qquad \text{(by (1231))}$$
$$= \det\begin{pmatrix} a & \ell & 0 & 0 \\ g & d & 0 & 0 \\ 0 & 0 & b & m \\ 0 & 0 & h & e \end{pmatrix} \cdot \det\begin{pmatrix} c & n \\ k & f \end{pmatrix} \qquad (1232)$$

(by Exercise 6.29, applied to $\begin{pmatrix} a & \ell & 0 & 0 \\ g & d & 0 & 0 \\ 0 & 0 & b & m \\ 0 & 0 & h & e \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} c & n \\ k & f \end{pmatrix}$, 2 and 4 instead of A, B, D, m and n).

But Exercise 6.29 (applied to $\begin{pmatrix} a & \ell \\ g & d \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} b & m \\ h & e \end{pmatrix}$, 2 and 2 instead of A, B, D, m and n) yields

$$\det\begin{pmatrix} a & \ell & 0 & 0 \\ g & d & 0 & 0 \\ 0 & 0 & b & m \\ 0 & 0 & h & e \end{pmatrix} = \underbrace{\det\begin{pmatrix} a & \ell \\ g & d \end{pmatrix}}_{=ad-\ell g} \cdot \underbrace{\det\begin{pmatrix} b & m \\ h & e \end{pmatrix}}_{=be-mh} = (ad-\ell g)\cdot(be-mh).$$

Hence, (1232) becomes

$$\det A = \underbrace{\det\begin{pmatrix} a & \ell & 0 & 0 \\ g & d & 0 & 0 \\ 0 & 0 & b & m \\ 0 & 0 & h & e \end{pmatrix}}_{=(ad-\ell g)\cdot(be-mh)} \cdot \underbrace{\det\begin{pmatrix} c & n \\ k & f \end{pmatrix}}_{=cf-nk} = (ad-\ell g)\cdot(be-mh)\cdot(cf-nk).$$

This solves Exercise 6.31 **(b)**. □

7.101. Solution to Exercise 6.33

Before we start solving this exercise, let us show some lemmas. The first of them is a (somewhat disguised) particular case of the Cauchy-Binet formula:

Lemma 7.180. Let n be a positive integer. Let A be an $(n-1) \times n$-matrix. Let B be an $n \times (n-1)$-matrix. Then,

$$\det(AB) = \sum_{k=1}^{n} \det\left(\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n} A\right) \cdot \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n} B\right).$$

Proof of Lemma 7.180. Let $[n]$ denote the set $\{1,2,\ldots,n\}$. Define a subset $\mathbf{I}$ of $[n]^{n-1}$ by

$$\mathbf{I} = \left\{(k_1,k_2,\ldots,k_{n-1}) \in [n]^{n-1} \mid k_1 < k_2 < \cdots < k_{n-1}\right\}.$$

Theorem 6.32 (applied to $n-1$ and n instead of n and m) yields

$$\det(AB) = \sum_{1 \leq g_1 < g_2 < \cdots < g_{n-1} \leq n} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_{n-1}} A\right) \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_{n-1}} B\right). \qquad (1233)$$

Recall that the summation sign $\sum\limits_{1 \leq g_1 < g_2 < \cdots < g_{n-1} \leq n-1}$ is an abbreviation for $\sum\limits_{\substack{(g_1,g_2,\ldots,g_{n-1}) \in \{1,2,\ldots,n\}^{n-1};\\ g_1<g_2<\cdots<g_{n-1}}}$, which can be rewritten as $\sum\limits_{(g_1,g_2,\ldots,g_{n-1}) \in \mathbf{I}}$ (because the $(n-1)$-tuples $(g_1,g_2,\ldots,g_{n-1}) \in \{1,2,\ldots,n\}^{n-1}$ satisfying $g_1 < g_2 < \cdots < g_{n-1}$ are precisely the elements of $\mathbf{I}$). Therefore, (1233) can be rewritten as

$$\det(AB) = \sum_{(g_1,g_2,\ldots,g_{n-1}) \in \mathbf{I}} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_{n-1}} A\right) \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_{n-1}} B\right). \qquad (1234)$$

Now, let us take a closer look at $\mathbf{I}$. The set $\mathbf{I}$ consists of all $(n-1)$-tuples $(k_1,k_2,\ldots,k_{n-1}) \in [n]^{n-1}$ satisfying $k_1 < k_2 < \cdots < k_{n-1}$. There are only n such $(n-1)$-tuples: namely, the $(n-1)$-tuples $\left(1,2,\ldots,\widehat{k},\ldots,n\right)$ for $k \in \{1,2,\ldots,n\}$. This is intuitively clear: If you want to choose an $(n-1)$-tuple $(k_1,k_2,\ldots,k_{n-1}) \in \mathbf{I}$, you can simply decide which of the n elements $1,2,\ldots,n$ you do **not** want to be an entry of $(k_1,k_2,\ldots,k_{n-1})$, and then the $(n-1)$-tuple $(k_1,k_2,\ldots,k_{n-1})$ will have to be the list of all the remaining $n-1$ elements of $\{1,2,\ldots,n\}$ in increasing order. Let us formalize this argument a bit more:

For every $k \in \{1,2,\ldots,n\}$, we have $\left(1,2,\ldots,\widehat{k},\ldots,n\right) \in \mathbf{I}$ (for obvious reasons). Hence, we can define a map

$$\Phi : \{1,2,\ldots,n\} \to \mathbf{I}$$

by

$$\left(\Phi(k) = \left(1,2,\ldots,\widehat{k},\ldots,n\right) \qquad \text{for every } k \in \{1,2,\ldots,n\}\right).$$

Consider this map Φ. This map Φ is injective[546] and surjective[547]. Hence, the map Φ is a bijection. In other words, the map

$$\{1,2,\ldots,n\} \to \mathbf{I}, \qquad k \mapsto \left(1,2,\ldots,\widehat{k},\ldots,n\right)$$

is a bijection (since this map is precisely Φ).

[546] *Proof.* Let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ be such that $\Phi(i) = \Phi(j)$. We shall show that $i = j$.

The definition of Φ yields $\Phi(i) = \left(1,2,\ldots,\widehat{i},\ldots,n\right)$. Hence, i is the only element of $\{1,2,\ldots,n\}$ that does not appear in $\Phi(i)$. Similarly, j is the only element of $\{1,2,\ldots,n\}$ that does not appear in $\Phi(j)$. In other words, j is the only element of $\{1,2,\ldots,n\}$ that does not appear in $\Phi(i)$ (since $\Phi(i) = \Phi(j)$). Comparing this with the fact that i is the only element of $\{1,2,\ldots,n\}$ that does not appear in $\Phi(i)$, we conclude that $i = j$.

Now, let us forget that we fixed i and j. We thus have proven that if $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ are such that $\Phi(i) = \Phi(j)$, then $i = j$. In other words, the map Φ is injective.

[547] *Proof.* Let $\mathbf{g} \in \mathbf{I}$. We shall show that $\mathbf{g} \in \Phi(\{1,2,\ldots,n\})$.

We have $\mathbf{g} \in \mathbf{I} = \left\{(k_1,k_2,\ldots,k_{n-1}) \in [n]^{n-1} \mid k_1 < k_2 < \cdots < k_{n-1}\right\}$. In other words, $\mathbf{g}$ can be written in the form $\mathbf{g} = (g_1,g_2,\ldots,g_{n-1})$ for some $(g_1,g_2,\ldots,g_{n-1}) \in [n]^{n-1}$ satisfying $g_1 < g_2 < \cdots < g_{n-1}$. Consider this $(g_1,g_2,\ldots,g_{n-1})$.

The integers $g_1,g_2,\ldots,g_{n-1}$ are distinct (since $g_1 < g_2 < \cdots < g_{n-1}$). Thus, $\{g_1,g_2,\ldots,g_{n-1}\}$ is an $(n-1)$-element subset of $[n]$. Therefore, its complement $[n] \setminus \{g_1,g_2,\ldots,g_{n-1}\}$ is a 1-element subset of $[n]$ (since $n-(n-1) = 1$). In other words, $[n] \setminus \{g_1,g_2,\ldots,g_{n-1}\} = \{k\}$ for some $k \in [n]$. Consider this k.

We have $k \in [n] = \{1,2,\ldots,n\}$. Since $\{g_1,g_2,\ldots,g_{n-1}\} \subseteq [n]$, we have

$$\{g_1,g_2,\ldots,g_{n-1}\} = [n] \setminus \underbrace{([n] \setminus \{g_1,g_2,\ldots,g_{n-1}\})}_{=\{k\}} = [n] \setminus \{k\}.$$

Now, recall that $g_1 < g_2 < \cdots < g_{n-1}$. Hence, the $(n-1)$-tuple $(g_1,g_2,\ldots,g_{n-1})$ is the list of all elements of the set $\{g_1,g_2,\ldots,g_{n-1}\}$ in increasing order. Since $\{g_1,g_2,\ldots,g_{n-1}\} = [n] \setminus \{k\}$, this rewrites as follows: The $(n-1)$-tuple $(g_1,g_2,\ldots,g_{n-1})$ is the list of all elements of the set $[n] \setminus \{k\}$ in increasing order. But clearly the latter list is $\left(1,2,\ldots,\widehat{k},\ldots,n\right)$. Thus, the $(n-1)$-tuple $(g_1,g_2,\ldots,g_{n-1})$ is the list $\left(1,2,\ldots,\widehat{k},\ldots,n\right)$. In other words, $(g_1,g_2,\ldots,g_{n-1}) = \left(1,2,\ldots,\widehat{k},\ldots,n\right)$, so that

$$\mathbf{g} = (g_1,g_2,\ldots,g_{n-1}) = \left(1,2,\ldots,\widehat{k},\ldots,n\right) = \Phi(k) \in \Phi(\{1,2,\ldots,n\}).$$

Now, let us forget that we fixed $\mathbf{g}$. We thus have proven that $\mathbf{g} \in \Phi(\{1,2,\ldots,n\})$ for every $\mathbf{g} \in \mathbf{I}$. In other words, $\mathbf{I} \subseteq \Phi(\{1,2,\ldots,n\})$. In other words, the map Φ is surjective.

Now, (1234) becomes

$$\det(AB) = \sum_{(g_1,g_2,\ldots,g_{n-1})\in\mathbf{I}} \det\left(\operatorname{cols}_{g_1,g_2,\ldots,g_{n-1}} A\right) \cdot \det\left(\operatorname{rows}_{g_1,g_2,\ldots,g_{n-1}} B\right)$$
$$= \underbrace{\sum_{k\in\{1,2,\ldots,n\}}}_{=\sum_{k=1}^{n}} \det\left(\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n} A\right) \cdot \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n} B\right)$$
$$\left(\begin{array}{c}\text{here, we have substituted } \left(1,2,\ldots,\widehat{k},\ldots,n\right) \text{ for}\\ (g_1,g_2,\ldots,g_{n-1}) \text{ in the sum, since the map}\\ \{1,2,\ldots,n\}\to\mathbf{I},\ k\mapsto\left(1,2,\ldots,\widehat{k},\ldots,n\right) \text{ is a bijection}\end{array}\right)$$
$$= \sum_{k=1}^{n} \det\left(\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n} A\right) \cdot \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n} B\right).$$

This proves Lemma 7.180. □

Here comes one more simple lemma:

Lemma 7.181. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let A be an $n \times p$-matrix. Let B be a $p \times m$-matrix. Let $i_1, i_2, \ldots, i_u$ be some elements of $\{1, 2, \ldots, n\}$. Let $j_1, j_2, \ldots, j_v$ be some elements of $\{1, 2, \ldots, m\}$. Then,

$$\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v}(AB) = \left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right) \cdot \left(\operatorname{cols}_{j_1,j_2,\ldots,j_v} B\right).$$

Proof of Lemma 7.181. Let us write the $n \times p$-matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p}$. Let us write the $p \times m$-matrix B in the form $B = \left(b_{i,j}\right)_{1\leq i\leq p,\ 1\leq j\leq m}$.

The definition of the product of two matrices yields $AB = \left(\sum_{k=1}^{p} a_{i,k}b_{k,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ (since $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p}$ and $B = \left(b_{i,j}\right)_{1\leq i\leq p,\ 1\leq j\leq m}$). Thus, the definition of $\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v}(AB)$ yields

$$\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v}(AB) = \left(\sum_{k=1}^{p} a_{i_x,k}b_{k,j_y}\right)_{1\leq x\leq u,\ 1\leq y\leq v}. \tag{1235}$$

On the other hand, $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p}$. Hence, the definition of $\operatorname{rows}_{i_1,i_2,\ldots,i_u} A$ yields

$$\operatorname{rows}_{i_1,i_2,\ldots,i_u} A = \left(a_{i_x,j}\right)_{1\leq x\leq u,\ 1\leq j\leq p} = \left(a_{i_i,j}\right)_{1\leq i\leq u,\ 1\leq j\leq p}$$

[548] (here, we renamed the index (x,j) as (i,j)).

Also, $B = (b_{i,j})_{1\leq i\leq p,\ 1\leq j\leq m}$. Hence, the definition of $\operatorname{cols}_{j_1,j_2,\ldots,j_v} B$ yields

$$\operatorname{cols}_{j_1,j_2,\ldots,j_v} B = \left(b_{i,j_y}\right)_{1\leq i\leq p,\ 1\leq y\leq v} = \left(b_{i,j_j}\right)_{1\leq i\leq p,\ 1\leq j\leq v}$$

(here, we renamed the index (i,y) as (i,j)).

We have $\operatorname{rows}_{i_1,i_2,\ldots,i_u} A = (a_{i_i,j})_{1\leq i\leq u,\ 1\leq j\leq p}$ and $\operatorname{cols}_{j_1,j_2,\ldots,j_v} B = \left(b_{i,j_j}\right)_{1\leq i\leq p,\ 1\leq j\leq v}$. The definition of the product of two matrices thus yields

$$\begin{aligned}\left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right)\cdot\left(\operatorname{cols}_{j_1,j_2,\ldots,j_v} B\right) &= \left(\sum_{k=1}^{p} a_{i_i,k} b_{k,j_j}\right)_{1\leq i\leq u,\ 1\leq j\leq v} \\ &= \left(\sum_{k=1}^{p} a_{i_x,k} b_{k,j_y}\right)_{1\leq x\leq u,\ 1\leq y\leq v}\end{aligned}$$

(here, we have renamed the index (i,j) as (x,y)). Comparing this with (1235), we obtain

$$\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_v}(AB) = \left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right)\cdot\left(\operatorname{cols}_{j_1,j_2,\ldots,j_v} B\right).$$

This proves Lemma 7.181. □

We note in passing that Lemma 7.181 leads to the following generalization of Theorem 6.32:

Corollary 7.182. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let A be an $n\times p$-matrix. Let B be a $p\times m$-matrix. Let $u \in \mathbb{N}$. Let $i_1, i_2, \ldots, i_u$ be some elements of $\{1,2,\ldots,n\}$. Let $j_1, j_2, \ldots, j_u$ be some elements of $\{1,2,\ldots,m\}$. Then,

$$\det\left(\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_u}(AB)\right) = \sum_{1\leq g_1<g_2<\cdots<g_u\leq p} \det\left(\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{g_1,g_2,\ldots,g_u} A\right)\cdot\det\left(\operatorname{sub}_{g_1,g_2,\ldots,g_u}^{j_1,j_2,\ldots,j_u} B\right).$$

(Here, the summation sign "$\sum\limits_{1\leq g_1<g_2<\cdots<g_u\leq p}$" has to be interpreted as "$\sum\limits_{\substack{(g_1,g_2,\ldots,g_u)\in\{1,2,\ldots,p\}^u;\\ g_1<g_2<\cdots<g_u}}$", in analogy to Remark 6.33.)

Corollary 7.182 is precisely the formula [NouYam02, (1.10)][549]. We shall not use Corollary 7.182 in what follows, but let us nevertheless prove it:

[548] The double use of the letter "i" in "i_i" might appear confusing. The first "i" is part of the notation i_k for $k \in \{1,2,\ldots,u\}$; the second "i" is an element of $\{1,2,\ldots,u\}$. These two "i"s are unrelated to each other. I hope the reader can easily tell them apart by the fact that the "i" that is part of the notation i_k always appears with a subscript, whereas the second "i" never does.

[549] We notice that the notation $A_{j_1,j_2,\ldots,j_u}^{i_1,i_2,\ldots,i_u}$ in [NouYam02] is equivalent to our notation $\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_u} A$.

Proof of Corollary 7.182. Fix any $(g_1, g_2, \ldots, g_u) \in \{1, 2, \ldots, p\}^u$. Applying Proposition 6.79 **(d)** to p, u and $(g_1, g_2, \ldots, g_u)$ instead of m, v and $(j_1, j_2, \ldots, j_v)$, we obtain

$$\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{g_1,g_2,\ldots,g_u} A = \operatorname{rows}_{i_1,i_2,\ldots,i_u} \left(\operatorname{cols}_{g_1,g_2,\ldots,g_u} A\right) = \operatorname{cols}_{g_1,g_2,\ldots,g_u} \left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right).$$

Applying Proposition 6.79 **(d)** to p, B, u and $(g_1, g_2, \ldots, g_u)$ instead of n, A, v and $(i_1, i_2, \ldots, i_u)$, we obtain

$$\operatorname{sub}_{g_1,g_2,\ldots,g_u}^{j_1,j_2,\ldots,j_u} B = \operatorname{rows}_{g_1,g_2,\ldots,g_u} \left(\operatorname{cols}_{j_1,j_2,\ldots,j_u} B\right) = \operatorname{cols}_{j_1,j_2,\ldots,j_u} \left(\operatorname{rows}_{g_1,g_2,\ldots,g_u} B\right).$$

Now, Lemma 7.181 (applied to $v = u$) shows that

$$\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_u} (AB) = \left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right) \cdot \left(\operatorname{cols}_{j_1,j_2,\ldots,j_u} B\right).$$

Hence,

$$\begin{aligned}
&\det\left(\underbrace{\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{j_1,j_2,\ldots,j_u} (AB)}_{=\left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right)\cdot\left(\operatorname{cols}_{j_1,j_2,\ldots,j_u} B\right)}\right) \\
&= \det\left(\left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right) \cdot \left(\operatorname{cols}_{j_1,j_2,\ldots,j_u} B\right)\right) \\
&= \sum_{1 \leq g_1 < g_2 < \cdots < g_u \leq p} \det\left(\underbrace{\operatorname{cols}_{g_1,g_2,\ldots,g_u} \left(\operatorname{rows}_{i_1,i_2,\ldots,i_u} A\right)}_{=\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{g_1,g_2,\ldots,g_u} A}\right) \\
&\qquad \cdot \det\left(\underbrace{\operatorname{rows}_{g_1,g_2,\ldots,g_u} \left(\operatorname{cols}_{j_1,j_2,\ldots,j_u} B\right)}_{=\operatorname{sub}_{g_1,g_2,\ldots,g_u}^{j_1,j_2,\ldots,j_u} B}\right) \\
&\qquad \left(\begin{array}{c}\text{by Theorem 6.32 (applied to } u,\ p,\ \operatorname{rows}_{i_1,i_2,\ldots,i_u} A \text{ and} \\ \operatorname{cols}_{j_1,j_2,\ldots,j_u} B \text{ instead of } n,\ m,\ A \text{ and } B)\end{array}\right) \\
&= \sum_{1 \leq g_1 < g_2 < \cdots < g_u \leq p} \det\left(\operatorname{sub}_{i_1,i_2,\ldots,i_u}^{g_1,g_2,\ldots,g_u} A\right) \cdot \det\left(\operatorname{sub}_{g_1,g_2,\ldots,g_u}^{j_1,j_2,\ldots,j_u} B\right).
\end{aligned}$$

This proves Corollary 7.182. □

Solution to Exercise 6.33. Let $(u, v) \in \{1, 2, \ldots, n\}^2$. Thus, $1 \leq u \leq n$, so that $n \geq 1$.

For every $k \in \{1, 2, \ldots, n\}$, we have

$$\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n} \left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right) = A_{\sim u,\sim k} \tag{1236}$$

[550] and

$$\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n}\left(\operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,n} B\right) = B_{\sim k,\sim v} \tag{1237}$$

[551].

We have

$$(AB)_{\sim u,\sim v} = \left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right) \cdot \left(\operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,n} B\right) \tag{1238}$$

[552]. Taking determinants on both sides of this equation, we obtain

$$\begin{aligned}
&\det\left((AB)_{\sim u,\sim v}\right) \\
&= \det\left(\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right) \cdot \left(\operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,n} B\right)\right) \\
&= \sum_{k=1}^{n} \det\left(\underbrace{\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n}\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right)}_{\substack{=A_{\sim u,\sim k} \\ \text{(by (1236))}}}\right) \cdot \det\left(\underbrace{\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n}\left(\operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,n} B\right)}_{\substack{=B_{\sim k,\sim v} \\ \text{(by (1237))}}}\right) \\
&\qquad \left(\begin{array}{c}\text{by Lemma 7.180, applied to } \operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A \\ \text{and } \operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,n} B \text{ instead of } A \text{ and } B\end{array}\right) \\
&= \sum_{k=1}^{n} \det\left(A_{\sim u,\sim k}\right) \cdot \det\left(B_{\sim k,\sim v}\right).
\end{aligned} \tag{1239}$$

Let us now forget that we fixed (u,v). We thus have proven (1239) for every $(u,v) \in \{1,2,\ldots,n\}^2$.

Now, the definition of $\operatorname{adj}(AB)$ yields

$$\operatorname{adj}(AB) = \left((-1)^{i+j} \det\left((AB)_{\sim j,\sim i}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}. \tag{1240}$$

---

[550] *Proof of (1236):* Let $k \in \{1,2,\ldots,n\}$. We can apply Proposition 6.79 **(d)** to $n-1$, $n-1$, $(1,2,\ldots,\widehat{u},\ldots,n)$ and $\left(1,2,\ldots,\widehat{k},\ldots,n\right)$ instead of u, v, $(i_1,i_2,\ldots,i_u)$ and $(j_1,j_2,\ldots,j_v)$. As a result, we obtain

$$\operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,n}_{1,2,\ldots,\widehat{u},\ldots,n} A = \operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n}\left(\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n} A\right) = \operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n}\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right).$$

But the definition of $A_{\sim u,\sim k}$ yields

$$A_{\sim u,\sim k} = \operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,n}_{1,2,\ldots,\widehat{u},\ldots,n} A = \operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n}\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right).$$

This proves (1236).

[551] This holds for similar reasons.

[552] *Proof of (1238):* The definition of $(AB)_{\sim u,\sim v}$ yields

$$(AB)_{\sim u,\sim v} = \operatorname{sub}^{1,2,\ldots,\widehat{v},\ldots,n}_{1,2,\ldots,\widehat{u},\ldots,n}(AB) = \left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,n} A\right) \cdot \left(\operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,n} B\right)$$

(by Lemma 7.181, applied to $m=n$, $p=n$, $u=n-1$, $(i_1,i_2,\ldots,i_u) = (1,2,\ldots,\widehat{u},\ldots,n)$ and $(j_1,j_2,\ldots,j_v) = (1,2,\ldots,\widehat{v},\ldots,n)$). This proves (1238).

But every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfies

$$\begin{aligned}
(-1)^{i+j} & \underbrace{\det\left((AB)_{\sim j,\sim i}\right)}_{\substack{=\sum\limits_{k=1}^{n} \det\left(A_{\sim j,\sim k}\right)\cdot\det\left(B_{\sim k,\sim i}\right) \\ \text{(by (1239), applied to } (u,v)=(j,i))}} \\
&= (-1)^{i+j} \sum_{k=1}^{n} \det\left(A_{\sim j,\sim k}\right) \cdot \det\left(B_{\sim k,\sim i}\right) \\
&= \sum_{k=1}^{n} \underbrace{(-1)^{i+j}}_{\substack{=(-1)^{(i+k)+(k+j)} \\ \text{(since } i+j\equiv i+j+2k=(i+k)+(k+j) \bmod 2)}} \det\left(A_{\sim j,\sim k}\right) \cdot \det\left(B_{\sim k,\sim i}\right) \\
&= \sum_{k=1}^{n} \underbrace{(-1)^{(i+k)+(k+j)}}_{=(-1)^{i+k}(-1)^{k+j}} \det\left(A_{\sim j,\sim k}\right) \cdot \det\left(B_{\sim k,\sim i}\right) \\
&= \sum_{k=1}^{n} (-1)^{i+k} \det\left(B_{\sim k,\sim i}\right) \cdot (-1)^{k+j} \det\left(A_{\sim j,\sim k}\right).
\end{aligned}$$

Thus, (1240) becomes

$$\begin{aligned}
\operatorname{adj}(AB) &= \left(\underbrace{(-1)^{i+j} \det\left((AB)_{\sim j,\sim i}\right)}_{=\sum\limits_{k=1}^{n} (-1)^{i+k} \det\left(B_{\sim k,\sim i}\right)\cdot(-1)^{k+j} \det\left(A_{\sim j,\sim k}\right)} \right)_{1\leq i\leq n,\ 1\leq j\leq n} \\
&= \left(\sum_{k=1}^{n} (-1)^{i+k} \det\left(B_{\sim k,\sim i}\right) \cdot (-1)^{k+j} \det\left(A_{\sim j,\sim k}\right) \right)_{1\leq i\leq n,\ 1\leq j\leq n} \qquad (1241)
\end{aligned}$$

On the other hand, we have $\operatorname{adj} B = \left((-1)^{i+j} \det\left(B_{\sim j,\sim i}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of $\operatorname{adj} B$) and $\operatorname{adj} A = \left((-1)^{i+j} \det\left(A_{\sim j,\sim i}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of $\operatorname{adj} A$). Therefore, the definition of the product of two matrices shows that

$$\operatorname{adj} B \cdot \operatorname{adj} A = \left(\sum_{k=1}^{n} (-1)^{i+k} \det\left(B_{\sim k,\sim i}\right) \cdot (-1)^{k+j} \det\left(A_{\sim j,\sim k}\right) \right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Compared with (1241), this yields $\operatorname{adj}(AB) = \operatorname{adj} B \cdot \operatorname{adj} A$. This solves Exercise 6.33. □

7.102. Solution to Exercise 6.34

Throughout this section, we shall use the notation $V(y_1, y_2, \ldots, y_n)$ defined in Exercise 6.34. Let us now make some preparations for solving Exercise 6.34.

7.102.1. Lemmas

Definition 7.183. If i and j are two objects, then $\delta_{i,j}$ is defined to be the element $\begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$ of $\mathbb{K}$.

Let us now prove four mostly trivial lemmas:

Lemma 7.184. Let $n \in \mathbb{N}$. Then,

$$\sum_{r=0}^{n-1} r = \binom{n}{2}. \qquad (1242)$$

Proof of Lemma 7.184. If $n = 0$, then Lemma 7.184 holds for obvious reasons. Thus, we WLOG assume that $n \neq 0$. Hence, $n \geq 1$, so that $n - 1 \in \mathbb{N}$. Now,

$$\begin{aligned} \sum_{r=0}^{n-1} r &= \sum_{i=0}^{n-1} i \qquad \text{(here, we have renamed the summation index } r \text{ as } i\text{)} \\ &= \frac{(n-1)((n-1)+1)}{2} \qquad \text{(by (13) (applied to } n-1 \text{ instead of } n\text{))} \\ &= \frac{n(n-1)}{2} = \binom{n}{2}. \end{aligned}$$

This proves Lemma 7.184. □

Lemma 7.185. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. Let $k \in \{1, 2, \ldots, n\}$. For each $j \in \{1, 2, \ldots, n\}$, set $y_j = x_j + \delta_{j,k} t$. Then,

$$(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n) = (y_1, y_2, \ldots, y_n).$$

Proof of Lemma 7.185. For each $j \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned} &(\text{the } j\text{-th entry of the list } (y_1, y_2, \ldots, y_n)) \\ &= y_j = x_j + \underbrace{\delta_{j,k}}_{=\begin{cases} 1, & \text{if } j = k; \\ 0, & \text{if } j \neq k \end{cases}} t = x_j + \begin{cases} 1, & \text{if } j = k; \\ 0, & \text{if } j \neq k \end{cases} t \\ &= \begin{cases} x_j + 1t, & \text{if } j = k; \\ x_j + 0t, & \text{if } j \neq k \end{cases} = \begin{cases} x_j + t, & \text{if } j = k; \\ x_j, & \text{if } j \neq k \end{cases} \\ &= \begin{cases} x_k + t, & \text{if } j = k; \\ x_j, & \text{if } j \neq k \end{cases} \qquad \left(\text{since } x_j = x_k \text{ in the case when } j = k\right) \\ &= (\text{the } j\text{-th entry of the list } (x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n)). \end{aligned}$$

Hence, $(y_1, y_2, \ldots, y_n) = (x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n)$. This proves Lemma 7.185. $\square$

Lemma 7.186. Let $n \in \mathbb{N}$. Let $y_1, y_2, \ldots, y_n$ be n elements of $\mathbb{K}$. Then,

$$V(y_1, y_2, \ldots, y_n) = \det\left(\left(y_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

Proof of Lemma 7.186. Theorem 6.46 **(a)** (applied to $x_i = y_i$) yields

$$\det\left(\left(y_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \prod_{1\leq i<j\leq n} (y_i - y_j).$$

Comparing this with

$$V(y_1, y_2, \ldots, y_n) = \prod_{1\leq i<j\leq n} (y_i - y_j) \qquad \text{(by the definition of } V(y_1, y_2, \ldots, y_n)),$$

we obtain $V(y_1, y_2, \ldots, y_n) = \det\left(\left(y_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$. Thus, Lemma 7.186 is proven. $\square$

Lemma 7.187. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $k \in \{1, 2, \ldots, n\}$. Let $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$ and $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$ be two $n \times m$-matrices. Assume that every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$ satisfying $i \neq k$ satisfy

$$a_{i,j} = b_{i,j}. \tag{1243}$$

Then, $B_{\sim k,\sim q} = A_{\sim k,\sim q}$ for every $q \in \{1, 2, \ldots, m\}$.

Proof of Lemma 7.187. Let $q \in \{1, 2, \ldots, m\}$. The condition (1243) shows that the matrices A and B agree in all their entries except for those in the respective k-th rows of the matrices. Therefore, the matrices $A_{\sim k,\sim q}$ and $B_{\sim k,\sim q}$ must agree completely (since they are obtained from the matrices A and B by removing the k-th rows and the q-th columns). In other words, $A_{\sim k,\sim q} = B_{\sim k,\sim q}$. This proves Lemma 7.187. $\square$

Next, let us combine Theorem 6.82 **(b)** and Proposition 6.96 **(b)** into a convenient package:

Lemma 7.188. Let $n \in \mathbb{N}$. Let $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n \times n$-matrix. Let $q \in \{1, 2, \ldots, n\}$ and $r \in \{1, 2, \ldots, n\}$. Then,

$$\sum_{p=1}^{n} (-1)^{p+q} a_{p,r} \det\left(A_{\sim p,\sim q}\right) = \delta_{q,r} \det A.$$

Proof of Lemma 7.188. We are in one of the following two cases:

Case 1: We have $q = r$.

Case 2: We have $q \neq r$.

Let us first consider Case 1. In this case, we have $q = r$. Hence, $\delta_{q,r} = 1$, so that

$$\underbrace{\delta_{q,r}}_{=1} \det A = \det A = \sum_{p=1}^{n} (-1)^{p+q} \underbrace{a_{p,q}}_{\substack{=a_{p,r} \\ (\text{since } q=r)}} \det\left(A_{\sim p,\sim q}\right) \qquad (\text{by Theorem 6.82 } \textbf{(b)})$$
$$= \sum_{p=1}^{n} (-1)^{p+q} a_{p,r} \det\left(A_{\sim p,\sim q}\right).$$

Hence, Lemma 7.188 is proven in Case 1.

Let us now consider Case 2. In this case, we have $q \neq r$. Hence, $\delta_{q,r} = 0$, so that

$$\underbrace{\delta_{q,r}}_{=0} \det A = 0 \det A = 0$$
$$= \sum_{p=1}^{n} (-1)^{p+q} a_{p,r} \det\left(A_{\sim p,\sim q}\right) \qquad (\text{by Proposition 6.96 } \textbf{(b)}).$$

Hence, Lemma 7.188 is proven in Case 2.

We have now proven Lemma 7.188 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, this shows that Lemma 7.188 always holds. □

Next, we show three crucial lemmas:

Lemma 7.189. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Then,

$$V\left(x_1, x_2, \ldots, x_n\right) = \det A. \tag{1244}$$

Proof of Lemma 7.189. The definition of A yields $A = \left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Lemma 7.186 (applied to $y_j = x_j$) yields

$$V\left(x_1, x_2, \ldots, x_n\right) = \det\left(\underbrace{\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}}_{=A}\right) = \det A.$$

This proves Lemma 7.189. □

Lemma 7.190. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Let $k \in \{1, 2, \ldots, n\}$.

(a) We have

$$V(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{n} (-1)^{k+q} x_k^{n-q} \det\left(A_{\sim k, \sim q}\right).$$

(b) We have

$$V(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n) \\ = \sum_{q=1}^{n} (-1)^{k+q} (x_k + t)^{n-q} \det\left(A_{\sim k, \sim q}\right).$$

(c) We have

$$V(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n) - V(x_1, x_2, \ldots, x_n) \\ = \sum_{q=1}^{n} \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^{\ell} (-1)^{k+q} x_k^{n-q-\ell} \det\left(A_{\sim k, \sim q}\right).$$

Proof of Lemma 7.190. The definition of A yields $A = \left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

(a) Lemma 7.189 yields

$$V(x_1, x_2, \ldots, x_n) = \det A = \sum_{q=1}^{n} (-1)^{k+q} x_k^{n-q} \det\left(A_{\sim k, \sim q}\right)$$

(by Theorem 6.82 **(a)**, applied to x_i^{n-j} and k instead of $a_{i,j}$ and p). This proves Lemma 7.190 **(a)**.

(b) For each $j \in \{1, 2, \ldots, n\}$, define an element $y_j \in \mathbb{K}$ by $y_j = x_j + \delta_{j,k} t$. Define an $n \times n$-matrix B by $B = \left(y_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

Every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfying $i \neq k$ satisfy $x_i^{n-j} = y_i^{n-j}$ [553]. Hence, Lemma 7.187 (applied to n, x_i^{n-j} and y_i^{n-j} instead of m, $a_{i,j}$ and $b_{i,j}$) yields that

$$B_{\sim k, \sim q} = A_{\sim k, \sim q} \qquad \text{for every } q \in \{1, 2, \ldots, n\}. \tag{1245}$$

Moreover, the definition of y_k satisfies $y_k = x_k + \underbrace{\delta_{k,k}}_{\substack{=1 \\ (\text{since } k=k)}} t = x_k + \underbrace{1t}_{=t} = x_k + t$.

[553] *Proof.* Let $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ be such that $i \neq k$. Then, the definition of y_i yields $y_i = x_i + \underbrace{\delta_{i,k}}_{\substack{=0 \\ (\text{since } i\neq k)}} t = x_i + \underbrace{0t}_{=0} = x_i$. Hence, $x_i = y_i$, and thus $x_i^{n-j} = y_i^{n-j}$. Qed.

Lemma 7.185 yields

$$\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) = \left(y_1, y_2, \ldots, y_n\right).$$

Hence,

$$\begin{aligned}
& V \underbrace{\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right)}_{=\left(y_1, y_2, \ldots, y_n\right)} \\
&= V\left(y_1, y_2, \ldots, y_n\right) = \det \left(\underbrace{\left(y_i^{n-j}\right)_{1 \le i \le n,\ 1 \le j \le n}}_{=B} \right) \qquad \text{(by Lemma 7.186)} \\
&= \det B = \sum_{q=1}^{n} (-1)^{k+q} \underbrace{y_k^{n-q}}_{\substack{=(x_k+t)^{n-q} \\ \text{(since } y_k = x_k + t\text{)}}} \det \left(\underbrace{B_{\sim k, \sim q}}_{\substack{=A_{\sim k, \sim q} \\ \text{(by (1245))}}} \right) \\
&\qquad \left(\begin{array}{c} \text{by Theorem 6.82 \textbf{(a)} (applied to } B\text{, } y_i^{n-j} \text{ and } k \\ \text{instead of } A\text{, } a_{i,j} \text{ and } p\text{)} \end{array} \right) \\
&= \sum_{q=1}^{n} (-1)^{k+q} \left(x_k + t\right)^{n-q} \det \left(A_{\sim k, \sim q}\right).
\end{aligned}$$

This proves Lemma 7.190 **(b)**.

(c) For every $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $m \in \mathbb{N}$, we have

$$(a+b)^m = \sum_{\ell=0}^{m} \binom{m}{\ell} a^\ell b^{m-\ell}. \tag{1246}$$

(Indeed, this is precisely the equality (338), with the variables n and k renamed as m and ℓ.) Now, every $q \in \{1, 2, \ldots, n\}$ satisfies

$$\left(x_k + t\right)^{n-q} - x_k^{n-q} = \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^\ell x_k^{n-q-\ell} \tag{1247}$$

[554].

[554] *Proof of (1247):* Let $q \in \{1, 2, \ldots, n\}$. Thus, $q \le n$.

Set $m = n - q$. Then, $m = n - q \in \mathbb{N}$ (since $q \le n$). The equality (1246) (applied to $a = t$ and

$b = x_k$) yields

$$(t + x_k)^m = \sum_{\ell=0}^{m} \binom{m}{\ell} t^\ell x_k^{m-\ell} = \underbrace{\binom{m}{0}}_{=1} \underbrace{t^0}_{=1} \underbrace{x_k^{m-0}}_{=x_k^m} + \sum_{\ell=1}^{m} \binom{m}{\ell} t^\ell x_k^{m-\ell}$$

(here, we have split off the addend for $\ell = 0$ from the sum)

$$= x_k^m + \sum_{\ell=1}^{m} \binom{m}{\ell} t^\ell x_k^{m-\ell}.$$

Subtracting x_k^m from both sides of this equality, we obtain

$$(t + x_k)^m - x_k^m = \sum_{\ell=1}^{m} \binom{m}{\ell} t^\ell x_k^{m-\ell}.$$

Since $t + x_k = x_k + t$, this rewrites as

$$(x_k + t)^m - x_k^m = \sum_{\ell=1}^{m} \binom{m}{\ell} t^\ell x_k^{m-\ell}.$$

Since $m = n - q$, this rewrites as

$$(x_k + t)^{n-q} - x_k^{n-q} = \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^\ell x_k^{n-q-\ell}.$$

This proves (1247).

But

$$\underbrace{V\left(x_1,x_2,\ldots,x_{k-1},x_k+t,x_{k+1},x_{k+2},\ldots,x_n\right)}_{\substack{=\sum_{q=1}^{n}(-1)^{k+q}(x_k+t)^{n-q}\det\left(A_{\sim k,\sim q}\right)\\ \text{(by Lemma 7.190 (b))}}} - \underbrace{V\left(x_1,x_2,\ldots,x_n\right)}_{\substack{=\sum_{q=1}^{n}(-1)^{k+q}x_k^{n-q}\det\left(A_{\sim k,\sim q}\right)\\ \text{(by Lemma 7.190 (a))}}}$$
$$= \sum_{q=1}^{n}(-1)^{k+q}(x_k+t)^{n-q}\det\left(A_{\sim k,\sim q}\right) - \sum_{q=1}^{n}(-1)^{k+q}x_k^{n-q}\det\left(A_{\sim k,\sim q}\right)$$
$$= \sum_{q=1}^{n}(-1)^{k+q}\underbrace{\left((x_k+t)^{n-q}\det\left(A_{\sim k,\sim q}\right) - x_k^{n-q}\det\left(A_{\sim k,\sim q}\right)\right)}_{=\left((x_k+t)^{n-q}-x_k^{n-q}\right)\det\left(A_{\sim k,\sim q}\right)}$$
$$= \sum_{q=1}^{n}(-1)^{k+q}\underbrace{\left((x_k+t)^{n-q}-x_k^{n-q}\right)}_{\substack{=\sum_{\ell=1}^{n-q}\binom{n-q}{\ell}t^{\ell}x_k^{n-q-\ell}\\ \text{(by (1247))}}}\det\left(A_{\sim k,\sim q}\right)$$
$$= \sum_{q=1}^{n}(-1)^{k+q}\left(\sum_{\ell=1}^{n-q}\binom{n-q}{\ell}t^{\ell}x_k^{n-q-\ell}\right)\det\left(A_{\sim k,\sim q}\right)$$
$$= \sum_{q=1}^{n}\sum_{\ell=1}^{n-q}\underbrace{(-1)^{k+q}\binom{n-q}{\ell}t^{\ell}}_{=\binom{n-q}{\ell}t^{\ell}(-1)^{k+q}}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right)$$
$$= \sum_{q=1}^{n}\sum_{\ell=1}^{n-q}\binom{n-q}{\ell}t^{\ell}(-1)^{k+q}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right).$$

This proves Lemma 7.190 **(c)**. □

Lemma 7.191. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Let $q \in \{1,2,\ldots,n\}$. Let $\ell \in \{1,2,\ldots,n-q\}$. Then,

$$\sum_{k=1}^{n}(-1)^{k+q}x_k^{n-q-\ell+1}\det\left(A_{\sim k,\sim q}\right) = \delta_{\ell,1}\det A.$$

Proof of Lemma 7.191. From $\ell \in \{1,2,\ldots,n-q\}$, we obtain $\ell \geq 1$ and $\ell \leq n-q$. Now, $q + \underbrace{\ell}_{\leq n-q} - 1 \leq q + (n-q) - 1 = n-1 \leq n$. Also, $q \in \{1,2,\ldots,n\}$, so that $q \geq 1$. Hence, $\underbrace{q}_{\geq 1} + \underbrace{\ell}_{\geq 1} - 1 \geq 1+1-1 = 1$. Combining this with $q+\ell-1 \leq n$, we obtain $q+\ell-1 \in \{1,2,\ldots,n\}$.

It is straightforward to see that $\delta_{q,q+\ell-1} = \delta_{\ell,1}$.

We have $A = \left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of A) and $q+\ell-1 \in \{1,2,\ldots,n\}$. Hence, Lemma 7.188 (applied to x_i^{n-j} and $q+\ell-1$ instead of $a_{i,j}$ and r) yields

$$\sum_{p=1}^{n} (-1)^{p+q} x_p^{n-(q+\ell-1)} \det\left(A_{\sim p,\sim q}\right) = \underbrace{\delta_{q,q+\ell-1}}_{=\delta_{\ell,1}} \det A = \delta_{\ell,1} \det A.$$

Thus,

$$\begin{aligned}
\delta_{\ell,1} \det A &= \sum_{p=1}^{n} (-1)^{p+q} \underbrace{x_p^{n-(q+\ell-1)}}_{\substack{=x_p^{n-q-\ell+1} \\ (\text{since } n-(q+\ell-1)=n-q-\ell+1)}} \det\left(A_{\sim p,\sim q}\right) \\
&= \sum_{p=1}^{n} (-1)^{p+q} x_p^{n-q-\ell+1} \det\left(A_{\sim p,\sim q}\right) = \sum_{k=1}^{n} (-1)^{k+q} x_k^{n-q-\ell+1} \det\left(A_{\sim k,\sim q}\right)
\end{aligned}$$

(here, we have renamed the summation index p as k). This proves Lemma 7.191. □

7.102.2. The solution

Solution to Exercise 6.34. Let A be the $n\times n$-matrix $\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

We have

$$\begin{aligned}
&\sum_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) - \sum_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_n\right) \\
&= \sum_{k=1}^{n} x_k \underbrace{\left(V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) - V\left(x_1, x_2, \ldots, x_n\right)\right)}_{\substack{=\sum_{q=1}^{n}\sum_{\ell=1}^{n-q}\binom{n-q}{\ell} t^{\ell}(-1)^{k+q} x_k^{n-q-\ell} \det\left(A_{\sim k,\sim q}\right) \\ \text{(by Lemma 7.190 \textbf{(c)})}}} \\
&= \sum_{k=1}^{n} x_k \sum_{q=1}^{n} \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^{\ell} (-1)^{k+q} x_k^{n-q-\ell} \det\left(A_{\sim k,\sim q}\right) \\
&= \sum_{q=1}^{n} \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^{\ell} \sum_{k=1}^{n} \underbrace{x_k (-1)^{k+q} x_k^{n-q-\ell}}_{=(-1)^{k+q} x_k^{n-q-\ell+1}} \det\left(A_{\sim k,\sim q}\right) \\
&= \sum_{q=1}^{n} \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^{\ell} \underbrace{\sum_{k=1}^{n} (-1)^{k+q} x_k^{n-q-\ell+1} \det\left(A_{\sim k,\sim q}\right)}_{\substack{=\delta_{\ell,1} \det A \\ \text{(by Lemma 7.191)}}} \\
&= \sum_{q=1}^{n} \sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^{\ell} \delta_{\ell,1} \det A = \sum_{q=1}^{n} \left(\sum_{\ell=1}^{n-q} \binom{n-q}{\ell} t^{\ell} \delta_{\ell,1} \right) \det A \\
&= \sum_{r=0}^{n-1} \left(\sum_{\ell=1}^{r} \binom{r}{\ell} t^{\ell} \delta_{\ell,1} \right) \det A \qquad (1248)
\end{aligned}$$

(here, we have substituted r for $n-q$ in the outer sum).

But every $r \in \{0, 1, \ldots, n-1\}$ satisfies

$$\sum_{\ell=1}^{r} \binom{r}{\ell} t^{\ell} \delta_{\ell,1} = rt \qquad (1249)$$

[555].

[555] *Proof of (1249):* Let $r \in \{0, 1, \ldots, n-1\}$. We must prove (1249).

We are in one of the following two cases:

Case 1: We have $r \neq 0$.

Case 2: We have $r = 0$.

Let us consider Case 1 first. In this case, we have $r \neq 0$. Hence, $r > 0$. Thus, $1 \in \{1, 2, \ldots, r\}$. Thus, we can split off the addend for $\ell = 1$ from the sum $\sum_{\ell=1}^{r} \binom{r}{\ell} t^{\ell} \delta_{\ell,1}$. As a result, we obtain

$$\sum_{\ell=1}^{r} \binom{r}{\ell} t^{\ell} \delta_{\ell,1} = \underbrace{\binom{r}{1}}_{=r} \underbrace{t^1}_{=t} \underbrace{\delta_{1,1}}_{=1} + \sum_{\ell=2}^{r} \binom{r}{\ell} t^{\ell} \underbrace{\delta_{\ell,1}}_{\substack{=0 \\ \text{(since } \ell \neq 1)}} = rt + \underbrace{\sum_{\ell=2}^{r} \binom{r}{\ell} t^{\ell} 0}_{=0} = rt.$$

Now, (1248) becomes

$$\sum_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) - \sum_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_n\right)$$
$$= \sum_{r=0}^{n-1} \underbrace{\left(\sum_{\ell=1}^{r} \binom{r}{\ell} t^{\ell} \delta_{\ell,1}\right)}_{\substack{=rt \\ \text{(by (1249))}}} \det A = \sum_{r=0}^{n-1} rt \det A$$
$$= \underbrace{\left(\sum_{r=0}^{n-1} r\right)}_{\substack{=\binom{n}{2} \\ \text{(by (1242))}}} t \underbrace{\det A}_{\substack{=V(x_1,x_2,\ldots,x_n) \\ \text{(by (1244))}}} = \binom{n}{2} tV\left(x_1, x_2, \ldots, x_n\right).$$

If we add $\sum\limits_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_n\right)$ to both sides of this equality, then we obtain

$$\sum_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right)$$
$$= \binom{n}{2} tV\left(x_1, x_2, \ldots, x_n\right) + \sum_{k=1}^{n} x_k V\left(x_1, x_2, \ldots, x_n\right)$$
$$= \left(\binom{n}{2} t + \sum_{k=1}^{n} x_k\right) V\left(x_1, x_2, \ldots, x_n\right).$$

This solves Exercise 6.34. □

7.102.3. Addendum: a simpler variant

Exercise 6.34 is now solved, but let me discuss one further fact, which is a variation on it. Namely, the following holds:

Proposition 7.192. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. Then,

$$\sum_{k=1}^{n} V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) = nV\left(x_1, x_2, \ldots, x_n\right).$$

We can prove this similarly to how we solved Exercise 6.34. Instead of Lemma 7.191, we use the following variant of this lemma:

Thus, (1249) is proven in Case 1.

Proving (1249) in Case 2 is straightforward and left to the reader.

We now have proven (1249) in each of the two Cases 1 and 2. Thus, (1249) always holds.

Lemma 7.193. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Let $q \in \{1, 2, \ldots, n\}$. Let $\ell \in \{1, 2, \ldots, n-q\}$. Then,

$$\sum_{k=1}^{n} (-1)^{k+q} x_k^{n-q-\ell} \det \left(A_{\sim k, \sim q}\right) = 0.$$

Proof of Lemma 7.193. Left to the reader. (Very similar to the above proof of Lemma 7.191.) □

Proof of Proposition 7.192. This proof is similar to our solution of Exercise 6.34, but a lot simpler. Again, the reader can fill in the details. □

7.102.4. Addendum: another sum of Vandermonde determinants

Here is one more result similar to Exercise 6.34 and Proposition 7.192:

Proposition 7.194. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $m \in \{0, 1, \ldots, n-1\}$. Let $t \in \mathbb{K}$. Then,

$$\sum_{k=1}^{n} x_k^m V \left(x_1, x_2, \ldots, x_{k-1}, t, x_{k+1}, x_{k+2}, \ldots, x_n\right) = t^m V \left(x_1, x_2, \ldots, x_n\right).$$

We have already built all the tools necessary for the proof of this proposition. We just need to repurpose a few of them:

Lemma 7.195. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Let $k \in \{1, 2, \ldots, n\}$. Then,

$$V \left(x_1, x_2, \ldots, x_{k-1}, t, x_{k+1}, x_{k+2}, \ldots, x_n\right) = \sum_{q=1}^{n} (-1)^{k+q} t^{n-q} \det \left(A_{\sim k, \sim q}\right).$$

Proof of Lemma 7.195. Lemma 7.190 **(b)** (applied to $t - x_k$ instead of t) yields

$$V \left(x_1, x_2, \ldots, x_{k-1}, x_k + (t - x_k), x_{k+1}, x_{k+2}, \ldots, x_n\right)$$
$$= \sum_{q=1}^{n} (-1)^{k+q} \left(x_k + (t - x_k)\right)^{n-q} \det \left(A_{\sim k, \sim q}\right).$$

Since $x_k + (t - x_k) = t$, this rewrites as

$$V \left(x_1, x_2, \ldots, x_{k-1}, t, x_{k+1}, x_{k+2}, \ldots, x_n\right) = \sum_{q=1}^{n} (-1)^{k+q} t^{n-q} \det \left(A_{\sim k, \sim q}\right).$$

This proves Lemma 7.195. □

Lemma 7.196. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Let $q \in \{1, 2, \ldots, n\}$ and $m \in \{0, 1, \ldots, n-1\}$. Then,

$$\sum_{k=1}^{n} x_k^m (-1)^{k+q} \det\left(A_{\sim k, \sim q}\right) = \delta_{q,n-m} \det A.$$

Proof of Lemma 7.196. From $q \in \{1, 2, \ldots, n\}$, we obtain $1 \leq q \leq n$. Hence, $1 \leq n$, so that $n \geq 1$ and therefore $n \in \{1, 2, \ldots, n\}$.

We have $A = \left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of A) and $n - m \in \{1, 2, \ldots, n\}$ (since $m \in \{0, 1, \ldots, n-1\}$). Hence, Lemma 7.188 (applied to x_i^{n-j} and $n - m$ instead of $a_{i,j}$ and r) yields

$$\sum_{p=1}^{n} (-1)^{p+q} x_p^{n-(n-m)} \det\left(A_{\sim p, \sim q}\right) = \delta_{q,n-m} \det A.$$

Thus,

$$\begin{aligned}
\delta_{q,n-m} \det A &= \sum_{p=1}^{n} (-1)^{p+q} \underbrace{x_p^{n-(n-m)}}_{\substack{=x_p^m \\ (\text{since } n-(n-m)=m)}} \det\left(A_{\sim p, \sim q}\right) \\
&= \sum_{p=1}^{n} \underbrace{(-1)^{p+q} x_p^m}_{=x_p^m (-1)^{p+q}} \det\left(A_{\sim p, \sim q}\right) \\
&= \sum_{p=1}^{n} x_p^m (-1)^{p+q} \det\left(A_{\sim p, \sim q}\right) = \sum_{k=1}^{n} x_k^m (-1)^{k+q} \det\left(A_{\sim k, \sim q}\right)
\end{aligned}$$

(here, we have renamed the summation index p as k). This proves Lemma 7.196. □

Proof of Proposition 7.194. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

From $m \in \{0, 1, \ldots, n-1\}$, we obtain $n - m \in \{1, 2, \ldots, n\}$. But

$$\begin{aligned}
&\sum_{k=1}^{n} x_k^m \underbrace{V\left(x_1, x_2, \ldots, x_{k-1}, t, x_{k+1}, x_{k+2}, \ldots, x_n\right)}_{\substack{=\sum_{q=1}^{n}(-1)^{k+q} t^{n-q} \det\left(A_{\sim k, \sim q}\right) \\ \text{(by Lemma 7.195)}}} \\
&= \sum_{k=1}^{n} \underbrace{x_k^m \sum_{q=1}^{n} (-1)^{k+q} t^{n-q} \det\left(A_{\sim k,\sim q}\right)}_{=\sum_{q=1}^{n} x_k^m (-1)^{k+q} t^{n-q} \det\left(A_{\sim k,\sim q}\right)} \\
&= \underbrace{\sum_{k=1}^{n} \sum_{q=1}^{n}}_{=\sum_{q=1}^{n} \sum_{k=1}^{n}} x_k^m (-1)^{k+q} t^{n-q} \det\left(A_{\sim k,\sim q}\right) \\
&= \sum_{q=1}^{n} \underbrace{\sum_{k=1}^{n} x_k^m (-1)^{k+q} t^{n-q} \det\left(A_{\sim k,\sim q}\right)}_{=t^{n-q} \sum_{k=1}^{n} x_k^m (-1)^{k+q} \det\left(A_{\sim k,\sim q}\right)} \\
&= \underbrace{\sum_{q=1}^{n}}_{=\sum_{q \in \{1,2,\ldots,n\}}} t^{n-q} \underbrace{\sum_{k=1}^{n} x_k^m (-1)^{k+q} \det\left(A_{\sim k,\sim q}\right)}_{\substack{=\delta_{q,n-m} \det A \\ \text{(by Lemma 7.196)}}} = \sum_{q \in \{1,2,\ldots,n\}} t^{n-q} \delta_{q,n-m} \det A \\
&= \sum_{\substack{q \in \{1,2,\ldots,n\}; \\ q \neq n-m}} t^{n-q} \underbrace{\delta_{q,n-m}}_{\substack{=0 \\ \text{(since } q \neq n-m)}} \det A + \underbrace{t^{n-(n-m)}}_{\substack{=t^m \\ \text{(since } n-(n-m)=m)}} \underbrace{\delta_{n-m,n-m}}_{\substack{=1 \\ \text{(since } n-m=n-m)}} \det A \\
&\qquad \left(\begin{array}{c} \text{here, we have split off the addend for } q = n-m \text{ from the sum} \\ \text{(since } n-m \in \{1, 2, \ldots, n\}\text{)} \end{array}\right) \\
&= \underbrace{\sum_{\substack{q \in \{1,2,\ldots,n\}; \\ q \neq n-m}} t^{n-q} 0 \det A}_{=0} + t^m \det A = t^m \underbrace{\det A}_{\substack{=V(x_1, x_2, \ldots, x_n) \\ \text{(by (1244))}}} \\
&= t^m V\left(x_1, x_2, \ldots, x_n\right).
\end{aligned}$$

This proves Proposition 7.194. □

7.102.5. Addendum: analogues involving products of all but one x_j

Let us finally prove a much more complicated analogue of Exercise 6.34 and Proposition 7.192. We shall use the following notations:

Definition 7.197. Let $n \in \mathbb{N}$. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. As usual, let $\mathcal{P}([n])$ denote the powerset of $[n]$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$.

(a) For every $j \in \mathbb{N}$, define an element $e_j(x_1, x_2, \ldots, x_n) \in \mathbb{K}$ by

$$e_j(x_1, x_2, \ldots, x_n) = \sum_{\substack{I \subseteq [n]; \\ |I| = j}} \prod_{i \in I} x_i.$$

(Here, as usual, the summation sign $\sum\limits_{\substack{I \subseteq [n]; \\ |I| = j}}$ means $\sum\limits_{\substack{I \in \mathcal{P}([n]); \\ |I| = j}}$.)

(b) For every $t \in \mathbb{K}$, define an element $z_t(x_1, x_2, \ldots, x_n) \in \mathbb{K}$ by

$$z_t(x_1, x_2, \ldots, x_n) = \sum_{j=0}^{n-1} e_{n-1-j}(x_1, x_2, \ldots, x_n) t^j.$$

Proposition 7.198. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. For each $i \in \{1, 2, \ldots, n\}$, set $y_i = \prod\limits_{\substack{j \in \{1, 2, \ldots, n\}; \\ j \neq i}} x_j$. Then,

$$\sum_{k=1}^{n} y_k V(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n)$$
$$= z_{-t}(x_1, x_2, \ldots, x_n) \cdot V(x_1, x_2, \ldots, x_n).$$

Before we start proving Proposition 7.198, let us explore a few properties of the elements defined in Definition 7.197:

Proposition 7.199. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$.

(a) We have

$$\prod_{i=1}^{n} (x_i + t) = \sum_{j=0}^{n} e_{n-j}(x_1, x_2, \ldots, x_n) t^j.$$

(b) We have

$$e_n(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} x_i.$$

(c) We have

$$t \cdot z_t(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} (x_i + t) - \prod_{i=1}^{n} x_i.$$

(d) Assume that the element t of $\mathbb{K}$ is invertible. Then,

$$z_t(x_1, x_2, \ldots, x_n) = \frac{\prod\limits_{i=1}^{n} (x_i + t) - \prod\limits_{i=1}^{n} x_i}{t}.$$

Proposition 7.199 **(a)** is one of several interconnected results known as *Vieta's formulas*. Proposition 7.199 **(d)** gives an alternative description of $z_t(x_1, x_2, \ldots, x_n)$ in the case when t is invertible.

Proof of Proposition 7.199. We shall use the notations $[n]$, $\mathcal{P}([n])$ and $\sum\limits_{\substack{I\subseteq[n];\\ |I|=j}}$ as in Definition 7.197.

We have $|[n]| = n$. Hence, every subset I of $[n]$ satisfies $|I| \in \{0, 1, \ldots, n\}$.

(a) Exercise 6.1 **(a)** (applied to $a_i = x_i$ and $b_i = t$) yields

$$\begin{aligned}
\prod_{i=1}^{n}(x_i + t) &= \underbrace{\sum_{I\subseteq[n]}}_{\substack{=\sum\limits_{j\in\{0,1,\ldots,n\}}\sum\limits_{\substack{I\subseteq[n];\\ |I|=j}} \\ \text{(since every subset } I \text{ of } [n] \\ \text{satisfies } |I|\in\{0,1,\ldots,n\})}} \left(\prod_{i\in I} x_i\right) \underbrace{\left(\prod_{i\in[n]\setminus I} t\right)}_{\substack{=t^{|[n]\setminus I|}=t^{|[n]|-|I|} \\ \text{(since } |[n]\setminus I|=|[n]|-|I| \\ \text{(because } I\subseteq[n]))}} \\
&= \underbrace{\sum_{j\in\{0,1,\ldots,n\}}}_{=\sum\limits_{j=0}^{n}} \sum_{\substack{I\subseteq[n];\\ |I|=j}} \left(\prod_{i\in I} x_i\right) \underbrace{t^{|[n]|-|I|}}_{\substack{=t^{n-j} \\ \text{(since } |[n]|=n \text{ and } |I|=j)}} \\
&= \sum_{j=0}^{n} \sum_{\substack{I\subseteq[n];\\ |I|=j}} \left(\prod_{i\in I} x_i\right) t^{n-j}.
\end{aligned}$$

Comparing this with

$$\begin{aligned}
\sum_{j=0}^{n} e_{n-j}(x_1, x_2, \ldots, x_n)\, t^j &= \sum_{j=0}^{n} \underbrace{e_{n-(n-j)}(x_1, x_2, \ldots, x_n)}_{\substack{=e_j(x_1,x_2,\ldots,x_n) \\ =\sum\limits_{\substack{I\subseteq[n];\\ |I|=j}} \prod\limits_{i\in I} x_i}} t^{n-j} \\
&\qquad \text{(here, we have substituted } n-j \text{ for } j \text{ in the sum)} \\
&= \sum_{j=0}^{n} \sum_{\substack{I\subseteq[n];\\ |I|=j}} \left(\prod_{i\in I} x_i\right) t^{n-j},
\end{aligned}$$

we obtain

$$\prod_{i=1}^{n}(x_i + t) = \sum_{j=0}^{n} e_{n-j}(x_1, x_2, \ldots, x_n)\, t^j.$$

This proves Proposition 7.199 **(a)**.

(b) If I is a subset of $[n]$, then we have the following logical equivalence:

$$(|I| = n) \iff (I = [n]) . \tag{1250}$$

(The proof of (1250) is obvious, since $|[n]| = n$.)
Now, the definition of $e_n(x_1, x_2, \ldots, x_n)$ yields

$$\begin{aligned} e_n(x_1, x_2, \ldots, x_n) &= \underbrace{\sum_{\substack{I \subseteq [n]; \\ |I| = n}}}_{\substack{= \sum\limits_{\substack{I \subseteq [n]; \\ I = [n]}} \\ \text{(by the equivalence (1250))}}} \prod_{i \in I} x_i = \sum_{\substack{I \subseteq [n]; \\ I = [n]}} \prod_{i \in I} x_i \\ &= \underbrace{\prod_{i \in [n]}}_{= \prod\limits_{i=1}^{n}} x_i \qquad (\text{since } [n] \text{ is a subset of } [n]) \\ &= \prod_{i=1}^{n} x_i . \end{aligned}$$

This proves Proposition 7.199 **(b)**.

(c) We have $0 \in \{0, 1, \ldots, n\}$ (since $n \in \mathbb{N}$). Proposition 7.199 **(a)** shows that

$$\begin{aligned} \prod_{i=1}^{n} (x_i + t) &= \sum_{j=0}^{n} e_{n-j}(x_1, x_2, \ldots, x_n) t^j \\ &= \underbrace{e_{n-0}(x_1, x_2, \ldots, x_n)}_{\substack{= e_n(x_1, x_2, \ldots, x_n) \\ = \prod\limits_{i=1}^{n} x_i \\ \text{(by Proposition 7.199 (b))}}} \underbrace{t^0}_{=1} + \underbrace{\sum_{j=1}^{n} e_{n-j}(x_1, x_2, \ldots, x_n) t^j}_{\substack{= \sum\limits_{j=0}^{n-1} e_{n-(j+1)}(x_1, x_2, \ldots, x_n) t^{j+1} \\ \text{(here, we have substituted } j+1 \\ \text{for } j \text{ in the sum)}}} \\ &\qquad \left(\begin{array}{c} \text{here, we have split off the addend for } j = 0 \text{ from the sum} \\ \text{(since } 0 \in \{0, 1, \ldots, n\} \text{)} \end{array} \right) \\ &= \prod_{i=1}^{n} x_i + \sum_{j=0}^{n-1} \underbrace{e_{n-(j+1)}(x_1, x_2, \ldots, x_n)}_{\substack{= e_{n-1-j}(x_1, x_2, \ldots, x_n) \\ \text{(since } n-(j+1) = n-1-j)}} \underbrace{t^{j+1}}_{= t t^j} \\ &= \prod_{i=1}^{n} x_i + \sum_{j=0}^{n-1} e_{n-1-j}(x_1, x_2, \ldots, x_n) t t^j . \end{aligned}$$

Solving this equality for $\sum_{j=0}^{n-1} e_{n-1-j}(x_1, x_2, \ldots, x_n) tt^j$, we obtain

$$\sum_{j=0}^{n-1} e_{n-1-j}(x_1, x_2, \ldots, x_n) tt^j = \prod_{i=1}^{n} (x_i + t) - \prod_{i=1}^{n} x_i. \qquad (1251)$$

But

$$t \cdot \underbrace{z_t(x_1, x_2, \ldots, x_n)}_{\substack{= \sum_{j=0}^{n-1} e_{n-1-j}(x_1,x_2,\ldots,x_n) t^j \\ \text{(by the definition of } z_t(x_1,x_2,\ldots,x_n))}}$$
$$= t \cdot \sum_{j=0}^{n-1} e_{n-1-j}(x_1, x_2, \ldots, x_n) t^j = \sum_{j=0}^{n-1} e_{n-1-j}(x_1, x_2, \ldots, x_n) tt^j$$
$$= \prod_{i=1}^{n} (x_i + t) - \prod_{i=1}^{n} x_i \qquad \text{(by (1251))}.$$

This proves Proposition 7.199 **(c)**.

(d) Proposition 7.199 **(c)** yields

$$t \cdot z_t(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} (x_i + t) - \prod_{i=1}^{n} x_i.$$

We can divide both sides of this equality by t (since t is invertible). As a result, we obtain

$$z_t(x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} (x_i + t) - \prod_{i=1}^{n} x_i}{t}.$$

This proves Proposition 7.199 **(d)**. □

Our next proposition is crucial in getting a grip on the elements y_k in Proposition 7.198:

Proposition 7.200. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. For each $i \in \{1, 2, \ldots, n\}$, set $y_i = \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq i}} x_j$.

Let $k \in \{1, 2, \ldots, n\}$. Then,

$$y_k = \sum_{j=0}^{n-1} (-1)^{n-1-j} e_j(x_1, x_2, \ldots, x_n) x_k^{n-1-j}.$$

Our proof of Proposition 7.200 will rely on a basic fact about sets:

Proposition 7.201. For every set T and every $j \in \mathbb{N}$, we let $\mathcal{P}_j(T)$ denote the set of all j-element subsets of T.

Let S be a set. Let $s \in S$. Let m be a positive integer. Then:

- We have $\mathcal{P}_m(S \setminus \{s\}) \subseteq \mathcal{P}_m(S)$.
- The map

$$\mathcal{P}_{m-1}(S \setminus \{s\}) \to \mathcal{P}_m(S) \setminus \mathcal{P}_m(S \setminus \{s\}), \\ U \mapsto U \cup \{s\}$$

is well-defined and a bijection.

Roughly speaking, Proposition 7.201 claims that if s is an element of a set S, and if m is a positive integer, then:

- all m-element subsets of $S \setminus \{s\}$ are m-element subsets of S as well;
- the m-element subsets of S which are **not** m-element subsets of $S \setminus \{s\}$ are in bijection with the $(m-1)$-element subsets of $S \setminus \{s\}$; this bijection sends an $(m-1)$-element subset U of $S \setminus \{s\}$ to the m-element subset $U \cup \{s\}$ of S.

Restated this way, Proposition 7.201 should be intuitively clear. A rigorous proof is not hard to give[556].

Proof of Proposition 7.200. We shall use the notations $[n]$, $\mathcal{P}([n])$ and $\sum\limits_{\substack{I \subseteq [n]; \\ |I|=j}}$ as in Definition 7.197. We shall furthermore use the notation $\mathcal{P}_j(T)$ from Proposition 7.201.

Let $K = [n] \setminus \{k\}$. Note that $k \in \{1, 2, \ldots, n\} = [n]$, so that $|[n] \setminus \{k\}| = \underbrace{|[n]|}_{=n} - 1 = n - 1$. From $K = [n] \setminus \{k\}$, we obtain $|K| = |[n] \setminus \{k\}| = n - 1$.

For every $j \in \mathbb{N}$, define an element $f_j \in \mathbb{K}$ by

$$f_j = \sum_{\substack{I \subseteq K; \\ |I|=j}} \prod_{i \in I} x_i.$$

[556] Here is an outline: First, show that the two maps

$$\mathcal{P}_{m-1}(S \setminus \{s\}) \to \mathcal{P}_m(S) \setminus \mathcal{P}_m(S \setminus \{s\}), \\ U \mapsto U \cup \{s\}$$

and

$$\mathcal{P}_m(S) \setminus \mathcal{P}_m(S \setminus \{s\}) \to \mathcal{P}_{m-1}(S \setminus \{s\}), \\ U \mapsto U \setminus \{s\}$$

are well-defined. Then, show that these maps are mutually inverse. Conclude that the first of these two maps is invertible, i.e., is a bijection.

Then,

$$f_0 = 1 \qquad (1252)$$

[557].

Next, we shall show that

$$f_{n-1} = y_k. \qquad (1253)$$

[*Proof of (1253):* We have $|K| = n-1$. Hence, there exists only one $(n-1)$-element subset I of K: namely, the set K itself. In other words, there exists only one subset I of K satisfying $|I| = n-1$: namely, the set K itself. Hence, the sum $\sum_{\substack{I \subseteq K; \\ |I|=n-1}} \prod_{i \in I} x_i$ has only one addend: namely, the addend for $I = K$. Thus, this sum simplifies to

$$\sum_{\substack{I \subseteq K; \\ |I|=n-1}} \prod_{i \in I} x_i = \prod_{i \in K} x_i = \prod_{j \in K} x_j = \underbrace{\prod_{j \in \{1,2,\ldots,n\} \setminus \{k\}}}_{= \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq k}}} x_j$$
$$\left(\text{since } K = \underbrace{[n]}_{=\{1,2,\ldots,n\}} \setminus \{k\} = \{1,2,\ldots,n\} \setminus \{k\} \right)$$
$$= \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq k}} x_j.$$

Now, the definition of f_{n-1} yields

$$f_{n-1} = \sum_{\substack{I \subseteq K; \\ |I|=n-1}} \prod_{i \in I} x_i = \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq k}} x_j.$$

Comparing this with

$$y_k = \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq k}} x_j \qquad \left(\text{by the definition of } y_k\right),$$

we find $f_{n-1} = y_k$. Thus, (1253) is proven.]

[557] *Proof of (1252):* There exists exactly one subset I of K satisfying $|I| = 0$: namely, the subset $\varnothing$. Hence, the sum $\sum_{\substack{I \subseteq K; \\ |I|=0}} \prod_{i \in I} x_i$ has only one addend: namely, the addend for $I = \varnothing$. Thus, this sum simplifies to $\sum_{\substack{I \subseteq K; \\ |I|=0}} \prod_{i \in I} x_i = \prod_{i \in \varnothing} x_i = (\text{empty product}) = 1$.

Now, the definition of f_0 yields $f_0 = \sum_{\substack{I \subseteq K; \\ |I|=0}} \prod_{i \in I} x_i = 1$. This proves (1252).

But

$$e_0(x_1, x_2, \ldots, x_n) = 1 \tag{1254}$$

[558].

We shall now show that

$$e_m(x_1, x_2, \ldots, x_n) = f_m + x_k f_{m-1} \tag{1255}$$

for every positive integer m.

[*Proof of (1255):* Let m be a positive integer. The definition of f_m yields

$$f_m = \underbrace{\sum_{\substack{I \subseteq K; \\ |I| = m}}}_{\substack{= \sum_{I \text{ is an } m\text{-element subset of } K} = \sum_{I \in \mathcal{P}_m(K)} \\ \text{(since the set of all } m\text{-element} \\ \text{subsets of } K \text{ is } \mathcal{P}_m(K))}} \prod_{i \in I} x_i = \sum_{I \in \mathcal{P}_m(K)} \prod_{i \in I} x_i. \tag{1256}$$

The same argument (but applied to $m-1$ instead of m) yields

$$f_{m-1} = \sum_{I \in \mathcal{P}_{m-1}(K)} \prod_{i \in I} x_i. \tag{1257}$$

Recall that $k \in [n]$. Hence, we can apply Proposition 7.201 to $[n]$ and k instead of S and s. As a result, we obtain the following two results:

- We have $\mathcal{P}_m([n] \setminus \{k\}) \subseteq \mathcal{P}_m([n])$.
- The map

$$\mathcal{P}_{m-1}([n] \setminus \{k\}) \to \mathcal{P}_m([n]) \setminus \mathcal{P}_m([n] \setminus \{k\}),$$
$$U \mapsto U \cup \{k\}$$

  is well-defined and a bijection.

Since $[n] \setminus \{k\} = K$, these two results can be rewritten as follows:

- We have $\mathcal{P}_m(K) \subseteq \mathcal{P}_m([n])$.

[558] *Proof of (1254):* There exists exactly one subset I of $[n]$ satisfying $|I| = 0$: namely, the subset $\varnothing$. Hence, the sum $\sum_{\substack{I \subseteq [n]; \\ |I|=0}} \prod_{i \in I} x_i$ has only one addend: namely, the addend for $I = \varnothing$. Thus, this sum simplifies to $\sum_{\substack{I \subseteq [n]; \\ |I|=0}} \prod_{i \in I} x_i = \prod_{i \in \varnothing} x_i = (\text{empty product}) = 1$.

Now, the definition of $e_0(x_1, x_2, \ldots, x_n)$ yields $e_0(x_1, x_2, \ldots, x_n) = \sum_{\substack{I \subseteq [n]; \\ |I|=0}} \prod_{i \in I} x_i = 1$. This proves (1254).

- The map
$$\mathcal{P}_{m-1}(K) \to \mathcal{P}_m([n]) \setminus \mathcal{P}_m(K), \qquad U \mapsto U \cup \{k\}$$
is well-defined and a bijection.

Furthermore, every $I \in \mathcal{P}_{m-1}(K)$ satisfies
$$\prod_{i \in I \cup \{k\}} x_i = x_k \prod_{i \in I} x_i \tag{1258}$$
[559]. Now,
$$\begin{aligned}
&\sum_{I \in \mathcal{P}_m([n]) \setminus \mathcal{P}_m(K)} \prod_{i \in I} x_i \\
&= \sum_{U \in \mathcal{P}_{m-1}(K)} \prod_{i \in U \cup \{k\}} x_i \\
&\qquad \left(\begin{array}{c}\text{here, we have substituted } U \cup \{k\} \text{ for } I \text{ in the sum, since} \\ \text{the map } \mathcal{P}_{m-1}(K) \to \mathcal{P}_m([n]) \setminus \mathcal{P}_m(K),\ U \mapsto U \cup \{k\} \\ \text{is a bijection}\end{array}\right) \\
&= \sum_{I \in \mathcal{P}_{m-1}(K)} \underbrace{\prod_{i \in I \cup \{k\}} x_i}_{\substack{= x_k \prod_{i \in I} x_i \\ \text{(by (1258))}}} \qquad \left(\begin{array}{c}\text{here, we have renamed the} \\ \text{summation index } U \text{ as } I\end{array}\right) \\
&= \sum_{I \in \mathcal{P}_{m-1}(K)} x_k \prod_{i \in I} x_i = x_k \sum_{I \in \mathcal{P}_{m-1}(K)} \prod_{i \in I} x_i.
\end{aligned} \tag{1259}$$

[559] *Proof of (1258):* Let $I \in \mathcal{P}_{m-1}(K)$. Thus, I is an $(m-1)$-element subset of K (since $\mathcal{P}_{m-1}(K)$ is defined as the set of all $(m-1)$-element subsets of K). In particular, I is a subset of K. Thus, $I \subseteq K = [n] \setminus \{k\}$. Thus, $k \notin I$. Hence, the sets I and $\{k\}$ are disjoint. Thus,
$$\prod_{i \in I \cup \{k\}} x_i = \left(\prod_{i \in I} x_i\right) \underbrace{\left(\prod_{i \in \{k\}} x_i\right)}_{= x_k} = \left(\prod_{i \in I} x_i\right) x_k = x_k \prod_{i \in I} x_i.$$
This proves (1258).

But the definition of $e_m(x_1, x_2, \ldots, x_n)$ yields

$$\begin{aligned}
e_m(x_1, x_2, \ldots, x_n) &= \underbrace{\sum_{\substack{I \subseteq [n];\\ |I| = m}}}_{\substack{= \sum_{I \text{ is an } m\text{-element subset of } [n]} = \sum_{I \in \mathcal{P}_m([n])} \\ \text{(since the set of all } m\text{-element} \\ \text{subsets of } [n] \text{ is } \mathcal{P}_m([n]))}} \prod_{i \in I} x_i = \sum_{I \in \mathcal{P}_m([n])} \prod_{i \in I} x_i \\
&= \underbrace{\sum_{\substack{I \in \mathcal{P}_m([n]);\\ I \in \mathcal{P}_m(K)}}}_{\substack{= \sum_{I \in \mathcal{P}_m(K)} \\ \text{(since } \mathcal{P}_m(K) \subseteq \mathcal{P}_m([n]))}} \prod_{i \in I} x_i + \underbrace{\sum_{\substack{I \in \mathcal{P}_m([n]);\\ I \notin \mathcal{P}_m(K)}}}_{= \sum_{I \in \mathcal{P}_m([n]) \setminus \mathcal{P}_m(K)}} \prod_{i \in I} x_i \\
&\qquad \left(\begin{array}{c} \text{since each } I \in \mathcal{P}_m([n]) \text{ satisfies either } I \in \mathcal{P}_m(K) \\ \text{or } I \notin \mathcal{P}_m(K) \text{ (but not both)} \end{array}\right) \\
&= \sum_{I \in \mathcal{P}_m(K)} \prod_{i \in I} x_i + \underbrace{\sum_{I \in \mathcal{P}_m([n]) \setminus \mathcal{P}_m(K)} \prod_{i \in I} x_i}_{\substack{= x_k \sum_{I \in \mathcal{P}_{m-1}(K)} \prod_{i \in I} x_i \\ \text{(by (1259))}}} \\
&= \underbrace{\sum_{I \in \mathcal{P}_m(K)} \prod_{i \in I} x_i}_{\substack{= f_m \\ \text{(by (1256))}}} + x_k \underbrace{\sum_{I \in \mathcal{P}_{m-1}(K)} \prod_{i \in I} x_i}_{\substack{= f_{m-1} \\ \text{(by (1257))}}} \\
&= f_m + x_k f_{m-1}.
\end{aligned}$$

This proves (1255).]

Recall that $k \in \{1, 2, \ldots, n\}$. Hence, $1 \leq k \leq n$, so that $n \geq 1$, so that $n - 1 \in \mathbb{N}$. Hence, $0 \in \{0, 1, \ldots, n-1\}$ and $n - 1 \in \{0, 1, \ldots, n-1\}$.

Now,

$$\sum_{j=0}^{n-1}(-1)^{n-1-j}e_j(x_1,x_2,\ldots,x_n)x_k^{n-1-j}$$

$$=\underbrace{(-1)^{n-1-0}}_{=(-1)^{n-1}}\underbrace{e_0(x_1,x_2,\ldots,x_n)}_{\substack{=1\\ \text{(by (1254))}}}\underbrace{x_k^{n-1-0}}_{=x_k^{n-1}}+\sum_{j=1}^{n-1}(-1)^{n-1-j}\underbrace{e_j(x_1,x_2,\ldots,x_n)}_{\substack{=f_j+x_kf_{j-1}\\ \text{(by (1255) (applied to } m=j))}}x_k^{n-1-j}$$

$$\left(\begin{array}{c}\text{here, we have split off the addend for } j=0 \text{ from the sum,}\\ \text{since } 0\in\{0,1,\ldots,n-1\}\end{array}\right)$$

$$=(-1)^{n-1}x_k^{n-1}+\underbrace{\sum_{j=1}^{n-1}(-1)^{n-1-j}\left(f_j+x_kf_{j-1}\right)x_k^{n-1-j}}_{=\sum_{j=1}^{n-1}(-1)^{n-1-j}f_jx_k^{n-1-j}+\sum_{j=1}^{n-1}(-1)^{n-1-j}x_kf_{j-1}x_k^{n-1-j}}$$

$$=(-1)^{n-1}x_k^{n-1}+\sum_{j=1}^{n-1}(-1)^{n-1-j}f_jx_k^{n-1-j}+\sum_{j=1}^{n-1}(-1)^{n-1-j}x_kf_{j-1}x_k^{n-1-j}. \tag{1260}$$

But

$$\sum_{j=1}^{n-1}(-1)^{n-1-j}\underbrace{x_kf_{j-1}}_{=f_{j-1}x_k}x_k^{n-1-j}$$

$$=\sum_{j=1}^{n-1}(-1)^{n-1-j}f_{j-1}\underbrace{x_kx_k^{n-1-j}}_{=x_k^{(n-1-j)+1}=x_k^{n-j}}=\sum_{j=1}^{n-1}(-1)^{n-1-j}f_{j-1}x_k^{n-j}$$

$$=\sum_{j=0}^{(n-1)-1}\underbrace{(-1)^{n-1-(j+1)}}_{\substack{=(-1)^{n-j}\\ \text{(since } n-1-(j+1)=n-j-2\equiv n-j \bmod 2)}}\underbrace{f_{(j+1)-1}}_{=f_j}\underbrace{x_k^{n-(j+1)}}_{\substack{=x_k^{n-1-j}\\ \text{(since } n-(j+1)=n-1-j)}}$$

$$\text{(here, we have substituted } j+1 \text{ for } j \text{ in the sum)}$$

$$=\sum_{j=0}^{(n-1)-1}\underbrace{(-1)^{n-j}}_{\substack{=-(-1)^{n-j-1}=-(-1)^{n-1-j}\\ \text{(since } n-j-1=n-1-j)}}f_jx_k^{n-1-j}=\sum_{j=0}^{(n-1)-1}\left(-(-1)^{n-1-j}\right)f_jx_k^{n-1-j}$$

$$=-\sum_{j=0}^{(n-1)-1}(-1)^{n-1-j}f_jx_k^{n-1-j}. \tag{1261}$$

Also,

$$\begin{aligned}
&\sum_{j=0}^{n-1} (-1)^{n-1-j} f_j x_k^{n-1-j} \\
&= \underbrace{(-1)^{n-1-0}}_{=(-1)^{n-1}} \underbrace{f_0}_{\substack{=1 \\ \text{(by (1252))}}} \underbrace{x_k^{n-1-0}}_{=x_k^{n-1}} + \sum_{j=1}^{n-1} (-1)^{n-1-j} f_j x_k^{n-1-j} \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } j=0 \text{ from the sum,} \\ \text{since } 0 \in \{0,1,\ldots,n-1\}\end{array}\right) \\
&= (-1)^{n-1} x_k^{n-1} + \sum_{j=1}^{n-1} (-1)^{n-1-j} f_j x_k^{n-1-j},
\end{aligned}$$

so that

$$\begin{aligned}
&(-1)^{n-1} x_k^{n-1} + \sum_{j=1}^{n-1} (-1)^{n-1-j} f_j x_k^{n-1-j} \\
&= \sum_{j=0}^{n-1} (-1)^{n-1-j} f_j x_k^{n-1-j} \\
&= \sum_{j=0}^{(n-1)-1} (-1)^{n-1-j} f_j x_k^{n-1-j} + \underbrace{(-1)^{n-1-(n-1)}}_{=(-1)^0=1} \underbrace{f_{n-1}}_{\substack{=y_k \\ \text{(by (1253))}}} \underbrace{x_k^{n-1-(n-1)}}_{=x_k^0=1} \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } j=n-1 \text{ from the sum,} \\ \text{since } n-1 \in \{0,1,\ldots,n-1\}\end{array}\right) \\
&= \sum_{j=0}^{(n-1)-1} (-1)^{n-1-j} f_j x_k^{n-1-j} + y_k. \qquad (1262)
\end{aligned}$$

Now, (1260) becomes

$$\begin{aligned}
&\sum_{j=0}^{n-1} (-1)^{n-1-j} e_j (x_1, x_2, \ldots, x_n) x_k^{n-1-j} \\
&= \underbrace{(-1)^{n-1} x_k^{n-1} + \sum_{j=1}^{n-1} (-1)^{n-1-j} f_j x_k^{n-1-j}}_{\substack{= \sum_{j=0}^{(n-1)-1} (-1)^{n-1-j} f_j x_k^{n-1-j} + y_k \\ \text{(by (1262))}}} + \underbrace{\sum_{j=1}^{n-1} (-1)^{n-1-j} x_k f_{j-1} x_k^{n-1-j}}_{\substack{= -\sum_{j=0}^{(n-1)-1} (-1)^{n-1-j} f_j x_k^{n-1-j} \\ \text{(by (1261))}}} \\
&= \sum_{j=0}^{(n-1)-1} (-1)^{n-1-j} f_j x_k^{n-1-j} + y_k + \left(-\sum_{j=0}^{(n-1)-1} (-1)^{n-1-j} f_j x_k^{n-1-j}\right) \\
&= y_k.
\end{aligned}$$

This proves Proposition 7.200. □

Lemma 7.202. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. For each $i \in \{1, 2, \ldots, n\}$, set $y_i = \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq i}} x_j$.

Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Let $q \in \{1, 2, \ldots, n\}$. For every $j \in \mathbb{N}$, define an element $\mathfrak{e}_j \in \mathbb{K}$ by $\mathfrak{e}_j = e_j\left(x_1, x_2, \ldots, x_n\right)$. Then,

$$\sum_{k=1}^{n} (-1)^{k+q} y_k \det\left(A_{\sim k, \sim q}\right) = (-1)^{n-q} \mathfrak{e}_{q-1} \det A.$$

Proof of Lemma 7.202. We have $A = \left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of A). Hence, Lemma 7.188 (applied to $a_{i,j} = x_i^{n-j}$) yields that

$$\sum_{p=1}^{n} (-1)^{p+q} x_p^{n-r} \det\left(A_{\sim p, \sim q}\right) = \delta_{q,r} \det A \tag{1263}$$

for each $r \in \{1, 2, \ldots, n\}$.

Let $j \in \{0, 1, \ldots, n-1\}$. Thus, $j+1 \in \{1, 2, \ldots, n\}$. Hence, (1263) (applied to $r = j+1$) yields

$$\sum_{p=1}^{n} (-1)^{p+q} x_p^{n-(j+1)} \det\left(A_{\sim p, \sim q}\right) = \delta_{q,j+1} \det A.$$

Hence,

$$\begin{aligned}
\delta_{q,j+1} \det A &= \sum_{p=1}^{n} (-1)^{p+q} \underbrace{x_p^{n-(j+1)}}_{\substack{= x_p^{n-1-j} \\ \text{(since } n-(j+1) = n-1-j\text{)}}} \det\left(A_{\sim p, \sim q}\right) \\
&= \sum_{p=1}^{n} (-1)^{p+q} x_p^{n-1-j} \det\left(A_{\sim p, \sim q}\right) \\
&= \sum_{k=1}^{n} (-1)^{k+q} x_k^{n-1-j} \det\left(A_{\sim k, \sim q}\right)
\end{aligned} \tag{1264}$$

(here, we have renamed the summation index p as k).

Now, forget that we fixed j. We thus have proven (1264) for every $j \in \{0, 1, \ldots, n-1\}$.

Every $k \in \mathbb{N}$ satisfies

$$y_k = \sum_{j=0}^{n-1} (-1)^{n-1-j} \underbrace{e_j(x_1, x_2, \ldots, x_n)}_{\substack{=\mathfrak{e}_j \\ \text{(since } \mathfrak{e}_j = e_j(x_1,x_2,\ldots,x_n) \\ \text{(by the definition of } \mathfrak{e}_j))}} x_k^{n-1-j} \qquad \text{(by Proposition 7.200)}$$
$$= \sum_{j=0}^{n-1} (-1)^{n-1-j} \mathfrak{e}_j x_k^{n-1-j}.$$

Thus,

$$\sum_{k=1}^{n} (-1)^{k+q} \underbrace{y_k}_{= \sum_{j=0}^{n-1} (-1)^{n-1-j} \mathfrak{e}_j x_k^{n-1-j}} \det(A_{\sim k, \sim q})$$
$$= \sum_{k=1}^{n} (-1)^{k+q} \left(\sum_{j=0}^{n-1} (-1)^{n-1-j} \mathfrak{e}_j x_k^{n-1-j} \right) \det(A_{\sim k, \sim q})$$
$$= \underbrace{\sum_{k=1}^{n} \sum_{j=0}^{n-1}}_{= \sum_{j=0}^{n-1} \sum_{k=1}^{n}} (-1)^{k+q} (-1)^{n-1-j} \mathfrak{e}_j x_k^{n-1-j} \det(A_{\sim k, \sim q})$$
$$= \sum_{j=0}^{n-1} \underbrace{\sum_{k=1}^{n} (-1)^{k+q} (-1)^{n-1-j} \mathfrak{e}_j x_k^{n-1-j} \det(A_{\sim k, \sim q})}_{= (-1)^{n-1-j} \mathfrak{e}_j \sum_{k=1}^{n} (-1)^{k+q} x_k^{n-1-j} \det(A_{\sim k, \sim q})}$$
$$= \sum_{j=0}^{n-1} (-1)^{n-1-j} \mathfrak{e}_j \underbrace{\sum_{k=1}^{n} (-1)^{k+q} x_k^{n-1-j} \det(A_{\sim k, \sim q})}_{\substack{= \delta_{q,j+1} \det A \\ \text{(by (1264))}}}$$
$$= \sum_{j=0}^{n-1} (-1)^{n-1-j} \mathfrak{e}_j \delta_{q,j+1} \det A = \underbrace{\sum_{j=1}^{n}}_{= \sum_{j \in \{1,2,\ldots,n\}}} \underbrace{(-1)^{n-1-(j-1)}}_{\substack{= (-1)^{n-j} \\ \text{(since } n-1-(j-1) = n-j)}} \mathfrak{e}_{j-1} \underbrace{\delta_{q,(j-1)+1}}_{\substack{= \delta_{q,j} \\ \text{(since } (j-1)+1 = j)}} \det A$$

(here, we have substituted $j-1$ for j in the sum)

$$= \sum_{j \in \{1,2,\ldots,n\}} (-1)^{n-j} \mathfrak{e}_{j-1} \delta_{q,j} \det A$$

$$= (-1)^{n-q} \mathfrak{e}_{q-1} \underbrace{\delta_{q,q}}_{\substack{=1 \\ (\text{since } q=q)}} \det A + \sum_{\substack{j\in\{1,2,\ldots,n\};\\ j\neq q}} (-1)^{n-j} \mathfrak{e}_{j-1} \underbrace{\delta_{q,j}}_{\substack{=0 \\ (\text{since } q\neq j \\ (\text{since } j\neq q))}} \det A$$

$$\left(\begin{array}{c}\text{here, we have split off the addend for } j=q \text{ from the sum} \\ (\text{since } q\in\{1,2,\ldots,n\})\end{array}\right)$$

$$= (-1)^{n-q} \mathfrak{e}_{q-1} \det A + \underbrace{\sum_{\substack{j\in\{1,2,\ldots,n\};\\ j\neq q}} (-1)^{n-j} \mathfrak{e}_{j-1} 0 \det A}_{=0} = (-1)^{n-q} \mathfrak{e}_{q-1} \det A.$$

This proves Lemma 7.202. □

Lemma 7.203. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. For each $i \in \{1, 2, \ldots, n\}$, set $y_i = \prod_{\substack{j\in\{1,2,\ldots,n\};\\ j\neq i}} x_j$.

Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Let $q \in \{1, 2, \ldots, n\}$ and $\ell \in \{0, 1, \ldots, n-q\}$. For every $j \in \mathbb{N}$, define an element $\mathfrak{e}_j \in \mathbb{K}$ by $\mathfrak{e}_j = e_j(x_1, x_2, \ldots, x_n)$. Then,

$$\sum_{k=1}^{n} (-1)^{k+q} y_k x_k^{n-q-\ell} \det\left(A_{\sim k,\sim q}\right) = \delta_{\ell,n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A.$$

Proof of Lemma 7.203. We are in one of the following two cases:

Case 1: We have $\ell = n - q$.

Case 2: We have $\ell \neq n - q$.

Let us first consider Case 1. In this case, we have $\ell = n-q$. Thus, $n - q - \underbrace{\ell}_{=n-q} = n - q - (n-q) = 0$. Now,

$$\sum_{k=1}^{n} (-1)^{k+q} y_k \underbrace{x_k^{n-q-\ell}}_{\substack{=x_k^0 \\ (\text{since } n-q-\ell=0)}} \det\left(A_{\sim k,\sim q}\right)$$
$$= \sum_{k=1}^{n} (-1)^{k+q} y_k \underbrace{x_k^0}_{=1} \det\left(A_{\sim k,\sim q}\right)$$
$$= \sum_{k=1}^{n} (-1)^{k+q} y_k \det\left(A_{\sim k,\sim q}\right) = (-1)^{n-q} \mathfrak{e}_{q-1} \det A$$

(by Lemma 7.202). Comparing this with

$$\underbrace{\delta_{\ell,n-q}}_{\substack{=1 \\ (\text{since } \ell=n-q)}} (-1)^{n-q} \mathfrak{e}_{q-1} \det A = (-1)^{n-q} \mathfrak{e}_{q-1} \det A,$$

we obtain

$$\sum_{k=1}^{n} (-1)^{k+q} y_k x_k^{n-q-\ell} \det \left(A_{\sim k, \sim q} \right) = \delta_{\ell, n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A.$$

Hence, Lemma 7.203 is proven in Case 1.

Let us now consider Case 2. In this case, we have $\ell \neq n-q$. Combining this with $\ell \leq n-q$ (since $\ell \in \{0, 1, \ldots, n-q\}$), we obtain $\ell < n-q$. Hence, $\ell \leq (n-q) - 1$ (since ℓ and $n-q$ are integers). In other words, $\ell + 1 \leq n-q$. But $\ell \geq 0$ (since $\ell \in \{0, 1, \ldots, n-q\}$), so that $\ell + 1 \geq 1$. Combining this with $\ell + 1 \leq n-q$, we obtain $\ell + 1 \in \{1, 2, \ldots, n-q\}$. Thus, Lemma 7.193 (applied to $\ell + 1$ instead of ℓ) yields

$$\sum_{k=1}^{n} (-1)^{k+q} x_k^{n-q-(\ell+1)} \det \left(A_{\sim k, \sim q} \right) = 0. \tag{1265}$$

But every $k \in \{1, 2, \ldots, n\}$ satisfies

$$y_k x_k^{n-q-\ell} = \mathfrak{e}_n x_k^{n-q-(\ell+1)} \tag{1266}$$

[560]. Now,

$$\begin{aligned}
&\sum_{k=1}^{n} (-1)^{k+q} \underbrace{y_k x_k^{n-q-\ell}}_{\substack{=\mathfrak{e}_n x_k^{n-q-(\ell+1)} \\ \text{(by (1266))}}} \det\left(A_{\sim k,\sim q}\right) \\
&= \sum_{k=1}^{n} (-1)^{k+q} \mathfrak{e}_n x_k^{n-q-(\ell+1)} \det\left(A_{\sim k,\sim q}\right) \\
&= \mathfrak{e}_n \underbrace{\sum_{k=1}^{n} (-1)^{k+q} x_k^{n-q-(\ell+1)} \det\left(A_{\sim k,\sim q}\right)}_{\substack{=0 \\ \text{(by (1265))}}} = \mathfrak{e}_n 0 = 0.
\end{aligned}$$

Comparing this with

$$\underbrace{\delta_{\ell,n-q}}_{\substack{=0 \\ \text{(since } \ell \neq n-q)}} (-1)^{n-q} \mathfrak{e}_{q-1} \det A = 0 (-1)^{n-q} \mathfrak{e}_{q-1} \det A = 0,$$

we obtain

$$\sum_{k=1}^{n} (-1)^{k+q} y_k x_k^{n-q-\ell} \det\left(A_{\sim k,\sim q}\right) = \delta_{\ell,n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A.$$

---

[560]*Proof of (1266):* Let $k \in \{1,2,\ldots,n\}$. We have $n - q - \underbrace{\ell}_{\leq (n-q)-1} \geq n - q - ((n-q)-1) = 1$. Thus, $x_k^{n-q-\ell} = x_k x_k^{(n-q-\ell)-1} = x_k x_k^{n-q-(\ell+1)}$ (since $(n-q-\ell)-1 = n-q-(\ell+1)$).

The definition of y_k yields $y_k = \prod_{\substack{j\in\{1,2,\ldots,n\}; \\ j\neq k}} x_j$. But the definition of $\mathfrak{e}_n$ yields $\mathfrak{e}_n = e_n(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} x_i$ (by Proposition 7.199 **(b)** (applied to $t = 0$)). Hence,

$$\begin{aligned}
\mathfrak{e}_n &= \prod_{i=1}^{n} x_i = \underbrace{\prod_{j=1}^{n}}_{=\prod_{j\in\{1,2,\ldots,n\}}} x_j \qquad \left(\begin{array}{c}\text{here, we have renamed the index } i \text{ as } j \\ \text{in the product}\end{array}\right) \\
&= \prod_{j\in\{1,2,\ldots,n\}} x_j = x_k \underbrace{\prod_{\substack{j\in\{1,2,\ldots,n\}; \\ j\neq k}} x_j}_{=y_k} \qquad \left(\begin{array}{c}\text{here, we have split off the factor} \\ \text{for } j = k \text{ from the product}\end{array}\right) \\
&= x_k y_k = y_k x_k. \qquad (1267)
\end{aligned}$$

Now,

$$y_k \underbrace{x_k^{n-q-\ell}}_{=x_k x_k^{n-q-(\ell+1)}} = \underbrace{y_k x_k}_{\substack{=\mathfrak{e}_n \\ \text{(by (1267))}}} x_k^{n-q-(\ell+1)} = \mathfrak{e}_n x_k^{n-q-(\ell+1)}.$$

This proves (1266).

Hence, Lemma 7.203 is proven in Case 2.

We have now proven Lemma 7.203 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, this shows that Lemma 7.203 always holds. □

Lemma 7.204. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Let $k \in \{1, 2, \ldots, n\}$. Then,

$$\begin{aligned} & V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) \\ & = \sum_{q=1}^{n} \sum_{\ell=0}^{n-q} \binom{n-q}{\ell} t^{\ell} (-1)^{k+q} x_k^{n-q-\ell} \det\left(A_{\sim k, \sim q}\right). \end{aligned}$$

Proof of Lemma 7.204. The definition of A yields $A = \left(x_i^{n-j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

For every $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $m \in \mathbb{N}$, we have

$$(a+b)^m = \sum_{\ell=0}^{m} \binom{m}{\ell} a^{\ell} b^{m-\ell}. \tag{1268}$$

(Indeed, this is precisely the statement of (338), with the variables n and k renamed as m and ℓ.) Now, every $q \in \{1, 2, \ldots, n\}$ satisfies

$$(x_k + t)^{n-q} = \sum_{\ell=0}^{n-q} \binom{n-q}{\ell} t^{\ell} x_k^{n-q-\ell} \tag{1269}$$

[561].

[561] *Proof of (1269):* Let $q \in \{1, 2, \ldots, n\}$. Thus, $q \leq n$.

Set $m = n - q$. Then, $m = n - q \in \mathbb{N}$ (since $q \leq n$). Applying (1268) to $a = t$ and $b = x_k$, we obtain

$$(t + x_k)^m = \sum_{\ell=0}^{m} \binom{m}{\ell} t^{\ell} x_k^{m-\ell}.$$

Since $t + x_k = x_k + t$, this rewrites as

$$(x_k + t)^m = \sum_{\ell=0}^{m} \binom{m}{\ell} t^{\ell} x_k^{m-\ell}.$$

Since $m = n - q$, this rewrites as

$$(x_k + t)^{n-q} = \sum_{\ell=0}^{n-q} \binom{n-q}{\ell} t^{\ell} x_k^{n-q-\ell}.$$

This proves (1269).

But

$$V\left(x_1,x_2,\ldots,x_{k-1},x_k+t,x_{k+1},x_{k+2},\ldots,x_n\right)$$
$$=\sum_{q=1}^{n}(-1)^{k+q}\underbrace{(x_k+t)^{n-q}}_{\substack{=\sum_{\ell=0}^{n-q}\binom{n-q}{\ell}t^\ell x_k^{n-q-\ell}\\ \text{(by (1269))}}}\det\left(A_{\sim k,\sim q}\right)\qquad\text{(by Lemma 7.190 \textbf{(b)})}$$
$$=\sum_{q=1}^{n}(-1)^{k+q}\left(\sum_{\ell=0}^{n-q}\binom{n-q}{\ell}t^\ell x_k^{n-q-\ell}\right)\det\left(A_{\sim k,\sim q}\right)$$
$$=\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}\underbrace{(-1)^{k+q}\binom{n-q}{\ell}t^\ell}_{=\binom{n-q}{\ell}t^\ell(-1)^{k+q}}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right)$$
$$=\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}\binom{n-q}{\ell}t^\ell(-1)^{k+q}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right).$$

This proves Lemma 7.204. □

Proof of Proposition 7.198. For every $j\in\mathbb{N}$, define an element $\mathfrak{e}_j\in\mathbb{K}$ by $\mathfrak{e}_j=e_j\left(x_1,x_2,\ldots,x_n\right)$.

Let A be the $n\times n$-matrix $\left(x_i^{n-j}\right)_{1\le i\le n,\ 1\le j\le n}$. We have

$$\sum_{k=1}^{n}y_k\underbrace{V\left(x_1,x_2,\ldots,x_{k-1},x_k+t,x_{k+1},x_{k+2},\ldots,x_n\right)}_{\substack{=\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}\binom{n-q}{\ell}t^\ell(-1)^{k+q}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right)\\ \text{(by Lemma 7.204)}}}$$
$$=\sum_{k=1}^{n}y_k\underbrace{\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}\binom{n-q}{\ell}t^\ell(-1)^{k+q}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right)}_{=\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}y_k\binom{n-q}{\ell}t^\ell(-1)^{k+q}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right)}$$
$$=\underbrace{\sum_{k=1}^{n}\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}}_{=\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}\sum_{k=1}^{n}}\underbrace{y_k\binom{n-q}{\ell}t^\ell(-1)^{k+q}}_{=\binom{n-q}{\ell}t^\ell(-1)^{k+q}y_k}x_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right)$$
$$=\sum_{q=1}^{n}\sum_{\ell=0}^{n-q}\sum_{k=1}^{n}\binom{n-q}{\ell}t^\ell(-1)^{k+q}y_kx_k^{n-q-\ell}\det\left(A_{\sim k,\sim q}\right).\qquad(1270)$$

Now, let $q \in \{1, 2, \ldots, n\}$. Thus, $n - q \in \{0, 1, \ldots, n-1\} \subseteq \mathbb{N}$. Hence, $n - q \in \{0, 1, \ldots, n-q\}$. Furthermore, (233) (applied to $m = n - q$) yields $\dbinom{n-q}{n-q} = 1$. Now,

$$\begin{aligned}
&\underbrace{\sum_{\ell=0}^{n-q} \sum_{k=1}^{n} \binom{n-q}{\ell} t^{\ell} (-1)^{k+q} y_k x_k^{n-q-\ell} \det\left(A_{\sim k, \sim q}\right)}_{= \binom{n-q}{\ell} t^{\ell} \sum_{k=1}^{n} (-1)^{k+q} y_k x_k^{n-q-\ell} \det\left(A_{\sim k, \sim q}\right)} \\
&= \sum_{\ell=0}^{n-q} \binom{n-q}{\ell} t^{\ell} \underbrace{\sum_{k=1}^{n} (-1)^{k+q} y_k x_k^{n-q-\ell} \det\left(A_{\sim k, \sim q}\right)}_{\substack{= \delta_{\ell, n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A \\ \text{(by Lemma 7.203)}}} \\
&= \sum_{\ell=0}^{n-q} \binom{n-q}{\ell} t^{\ell} \delta_{\ell, n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A \\
&= \sum_{\ell=0}^{(n-q)-1} \binom{n-q}{\ell} t^{\ell} \underbrace{\delta_{\ell, n-q}}_{\substack{=0 \\ \text{(because } \ell \neq n-q \\ \text{(since } \ell \leq (n-q)-1 < n-q\text{))}}} (-1)^{n-q} \mathfrak{e}_{q-1} \det A \\
&\qquad + \underbrace{\binom{n-q}{n-q}}_{=1} t^{n-q} \underbrace{\delta_{n-q, n-q}}_{\substack{=1 \\ \text{(since } n-q=n-q\text{)}}} (-1)^{n-q} \mathfrak{e}_{q-1} \det A \\
&\qquad \left(\begin{array}{c} \text{here, we have split off the addend for } \ell = n-q \text{ from the sum} \\ \text{(since } n-q \in \{0, 1, \ldots, n-q\}\text{)} \end{array}\right) \\
&= \underbrace{\sum_{\ell=0}^{(n-q)-1} \binom{n-q}{\ell} t^{\ell} 0 (-1)^{n-q} \mathfrak{e}_{q-1} \det A}_{=0} + t^{n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A \\
&= t^{n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A.
\end{aligned} \tag{1271}$$

Now, forget that we fixed q. We thus have proven (1271) for every $q \in \{1, 2, \ldots, n\}$.

Now, (1270) becomes

$$\begin{aligned}
&\sum_{k=1}^{n} y_k V\left(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n\right) \\
&= \sum_{q=1}^{n} \underbrace{\sum_{\ell=0}^{n-q} \sum_{k=1}^{n} \binom{n-q}{\ell} t^{\ell} (-1)^{k+q} y_k x_k^{n-q-\ell} \det\left(A_{\sim k, \sim q}\right)}_{\substack{= t^{n-q}(-1)^{n-q} \mathfrak{e}_{q-1} \det A \\ \text{(by (1271))}}} \\
&= \sum_{q=1}^{n} t^{n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A. \qquad (1272)
\end{aligned}$$

But

$$\begin{aligned}
&\sum_{q=1}^{n} \underbrace{t^{n-q} (-1)^{n-q}}_{=(t\cdot(-1))^{n-q}} \mathfrak{e}_{q-1} \det A \\
&= \sum_{q=1}^{n} \Big(\underbrace{t \cdot (-1)}_{=-t} \Big)^{n-q} \mathfrak{e}_{q-1} \det A = \sum_{q=1}^{n} (-t)^{n-q} \mathfrak{e}_{q-1} \det A \\
&= \sum_{j=0}^{n-1} \underbrace{(-t)^{n-(n-j)}}_{\substack{=(-t)^j \\ \text{(since } n-(n-j)=j)}} \underbrace{\mathfrak{e}_{(n-j)-1}}_{\substack{=\mathfrak{e}_{n-1-j} \\ \text{(since } (n-j)-1=n-1-j)}} \det A \\
&\qquad \left(\text{here, we have substituted } n-j \text{ for } q \text{ in the sum}\right) \\
&= \sum_{j=0}^{n-1} (-t)^j \underbrace{\mathfrak{e}_{n-1-j}}_{\substack{=e_{n-1-j}(x_1,x_2,\ldots,x_n) \\ \text{(by the definition of } \mathfrak{e}_{n-1-j})}} \det A = \sum_{j=0}^{n-1} (-t)^j e_{n-1-j}\left(x_1, x_2, \ldots, x_n\right) \det A \\
&= \left(\sum_{j=0}^{n-1} (-t)^j e_{n-1-j}\left(x_1, x_2, \ldots, x_n\right) \right) \det A. \qquad (1273)
\end{aligned}$$

But the definition of $z_{-t}\left(x_1, x_2, \ldots, x_n\right)$ yields

$$\begin{aligned}
z_{-t}\left(x_1, x_2, \ldots, x_n\right) &= \sum_{j=0}^{n-1} e_{n-1-j}\left(x_1, x_2, \ldots, x_n\right) (-t)^j \\
&= \sum_{j=0}^{n-1} (-t)^j e_{n-1-j}\left(x_1, x_2, \ldots, x_n\right). \qquad (1274)
\end{aligned}$$

Thus, (1273) becomes

$$\sum_{q=1}^{n} t^{n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A$$
$$= \underbrace{\left(\sum_{j=0}^{n-1} (-t)^j e_{n-1-j} (x_1, x_2, \ldots, x_n) \right)}_{\substack{= z_{-t}(x_1, x_2, \ldots, x_n) \\ \text{(by (1274))}}} \underbrace{\det A}_{\substack{= V(x_1, x_2, \ldots, x_n) \\ \text{(by (1244))}}}$$
$$= z_{-t} (x_1, x_2, \ldots, x_n) \cdot V (x_1, x_2, \ldots, x_n) . \tag{1275}$$

Hence, (1272) becomes

$$\sum_{k=1}^{n} y_k V (x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, x_{k+2}, \ldots, x_n)$$
$$= \sum_{q=1}^{n} t^{n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A = z_{-t} (x_1, x_2, \ldots, x_n) \cdot V (x_1, x_2, \ldots, x_n)$$

(by (1275)). Thus, Proposition 7.198 is finally proven. □

Let us finish with one further identity, which is reminiscent of both Proposition 7.198 and Proposition 7.194:

Proposition 7.205. Let $n \in \mathbb{N}$. Let $x_1, x_2, \ldots, x_n$ be n elements of $\mathbb{K}$. Let $t \in \mathbb{K}$. For each $i \in \{1, 2, \ldots, n\}$, set $y_i = \prod_{\substack{j \in \{1,2,\ldots,n\}; \\ j \neq i}} x_j$. Then,

$$\sum_{k=1}^{n} y_k V (x_1, x_2, \ldots, x_{k-1}, t, x_{k+1}, x_{k+2}, \ldots, x_n)$$
$$= z_{-t} (x_1, x_2, \ldots, x_n) \cdot V (x_1, x_2, \ldots, x_n) .$$

The proof proceeds similarly as our proof of Proposition 7.194:

Proof of Proposition 7.205. For every $j \in \mathbb{N}$, define an element $\mathfrak{e}_j \in \mathbb{K}$ by $\mathfrak{e}_j = e_j (x_1, x_2, \ldots, x_n)$.

Let A be the $n \times n$-matrix $\left(x_i^{n-j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Now,

$$\begin{aligned}
&\sum_{k=1}^{n} y_k \underbrace{V\left(x_1, x_2, \ldots, x_{k-1}, t, x_{k+1}, x_{k+2}, \ldots, x_n\right)}_{\substack{=\sum_{q=1}^{n}(-1)^{k+q}t^{n-q}\det\left(A_{\sim k,\sim q}\right) \\ \text{(by Lemma 7.195)}}} \\
&= \sum_{k=1}^{n} \underbrace{y_k \sum_{q=1}^{n} (-1)^{k+q} t^{n-q} \det\left(A_{\sim k,\sim q}\right)}_{=\sum_{q=1}^{n} y_k(-1)^{k+q}t^{n-q}\det\left(A_{\sim k,\sim q}\right)} \\
&= \underbrace{\sum_{k=1}^{n}\sum_{q=1}^{n}}_{=\sum_{q=1}^{n}\sum_{k=1}^{n}} \underbrace{y_k (-1)^{k+q}}_{=(-1)^{k+q}y_k} t^{n-q} \det\left(A_{\sim k,\sim q}\right) \\
&= \sum_{q=1}^{n} \underbrace{\sum_{k=1}^{n} (-1)^{k+q} y_k t^{n-q} \det\left(A_{\sim k,\sim q}\right)}_{=t^{n-q}\sum_{k=1}^{n}(-1)^{k+q}y_k\det\left(A_{\sim k,\sim q}\right)} \\
&= \sum_{q=1}^{n} t^{n-q} \underbrace{\sum_{k=1}^{n} (-1)^{k+q} y_k \det\left(A_{\sim k,\sim q}\right)}_{\substack{=(-1)^{n-q}\mathfrak{e}_{q-1}\det A \\ \text{(by Lemma 7.202)}}} \\
&= \sum_{q=1}^{n} t^{n-q} (-1)^{n-q} \mathfrak{e}_{q-1} \det A = z_{-t}\left(x_1, x_2, \ldots, x_n\right) \cdot V\left(x_1, x_2, \ldots, x_n\right)
\end{aligned}$$

(by (1275)). This proves Proposition 7.205. □

7.103. Solution to Exercise 6.35

Solution to Exercise 6.35. Note that VB and WD are $m \times m$-matrices. Hence, $VB + WD$ is an $m \times m$-matrix. Also, $\begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix}$ and $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ are $(n+m) \times (n+m)$-matrices.

Exercise 6.28 (applied to n, m, n, m, n, m, I_n, $0_{n\times m}$, V, W, A, B, C and D instead of n, n', m, m', ℓ, ℓ', A, B, C, D, A', B', C' and D') yields

$$\begin{aligned}
\begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \begin{pmatrix} I_n A + 0_{n\times m} C & I_n B + 0_{n\times m} D \\ VA + WC & VB + WD \end{pmatrix} \\
&= \begin{pmatrix} A & B \\ 0_{m\times n} & VB + WD \end{pmatrix}
\end{aligned}$$

(since $I_n A + \underbrace{0_{n\times m} C}_{=0_{n\times n}} = I_n A = A$, $I_n B + \underbrace{0_{n\times m} D}_{=0_{n\times m}} = I_n B = B$ and $\underbrace{VA}_{=-WC} + WC = -WC + WC = 0_{m\times n}$). Taking determinants on both sides of this equality, we obtain

$$\det\left(\begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right) = \det\begin{pmatrix} A & B \\ 0_{m\times n} & VB+WD \end{pmatrix}$$
$$= \det A \cdot \det\left(VB+WD\right)$$

(by Exercise 6.29 (applied to $VB + WD$ instead of D)). Hence,

$$\begin{aligned}
&\det A \cdot \det\left(VB+WD\right) \\
&= \det\left(\begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right) \\
&= \underbrace{\det\begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix}}_{\substack{=\det(I_n)\cdot\det W \\ \text{(by Exercise 6.30 (applied} \\ \text{to } I_n,\ V \text{ and } W \text{ instead of } A,\ C \text{ and } D))}} \cdot \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} \\
&\qquad \left(\begin{array}{c}\text{by Theorem 6.23 (applied to } n+m,\ \begin{pmatrix} I_n & 0_{n\times m} \\ V & W \end{pmatrix} \\ \text{and } \begin{pmatrix} A & B \\ C & D \end{pmatrix} \text{ instead of } n,\ A \text{ and } B)\end{array}\right) \\
&= \underbrace{\det\left(I_n\right)}_{=1} \cdot \det W \cdot \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det W \cdot \det\begin{pmatrix} A & B \\ C & D \end{pmatrix}.
\end{aligned}$$

This solves Exercise 6.35. □

7.104. Solution to Exercise 6.36

Solution to Exercise 6.36. Recall that $\det\left(I_n\right) = 1$. The same argument (applied to m instead of n) yields $\det\left(I_m\right) = 1$.

The matrix A is invertible; thus, it has an inverse $A^{-1} \in \mathbb{K}^{n\times n}$. The matrices $I_m \in \mathbb{K}^{m\times m}$ and $-CA^{-1} \in \mathbb{K}^{m\times n}$ satisfy $-\left(CA^{-1}\right)A = -I_m C$ (since $-\left(CA^{-1}\right)A = -C\underbrace{A^{-1}A}_{=I_n} = -CI_n = -\underbrace{C}_{=I_m C} = -I_m C$). Hence, Exercise 6.35 (applied to $W = I_m$ and $V = -CA^{-1}$) yields

$$\det\left(I_m\right) \cdot \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \cdot \det\left(-CA^{-1}B + \underbrace{I_m D}_{=D}\right)$$
$$= \det A \cdot \det\left(\underbrace{-CA^{-1}B + D}_{=D-CA^{-1}B}\right) = \det A \cdot \det\left(D - CA^{-1}B\right).$$

Hence,

$$\det A \cdot \det \left(D - CA^{-1}B\right) = \underbrace{\det \left(I_m\right)}_{=1} \cdot \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

This solves Exercise 6.36. □

7.105. Solution to Exercise 6.37

Before we solve Exercise 6.37, let us show two really simple facts:

Lemma 7.206. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then,

$$\begin{pmatrix} I_n & 0_{n\times m} \\ 0_{m\times n} & I_m \end{pmatrix} = I_{n+m}.$$

Proof of Lemma 7.206. Easy, and left to the reader. □

Lemma 7.207. Let n, n', m and m' be four nonnegative integers.

Let $A \in \mathbb{K}^{n\times m}$ and $A' \in \mathbb{K}^{n\times m}$. Let $B \in \mathbb{K}^{n\times m'}$ and $B' \in \mathbb{K}^{n\times m'}$. Let $C \in \mathbb{K}^{n'\times m}$ and $C' \in \mathbb{K}^{n'\times m}$. Let $D \in \mathbb{K}^{n'\times m'}$ and $D' \in \mathbb{K}^{n'\times m'}$. Assume that $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$. Then, $A = A'$, $B = B'$, $C = C'$ and $D = D'$.

Proof of Lemma 7.207. Easy, and left to the reader. □

Solution to Exercise 6.37. The matrix $\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$ is the inverse of the matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$. In other words, we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = I_{n+m} \qquad \text{and}$$
$$\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = I_{n+m}.$$

But Lemma 7.206 yields

$$\begin{pmatrix} I_n & 0_{n\times m} \\ 0_{m\times n} & I_m \end{pmatrix} = I_{n+m} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A'A + B'C & A'B + B'D \\ C'A + D'C & C'B + D'D \end{pmatrix}$$

(by Exercise 6.28 (applied to n, m, n, m, n, m, A', B', C', D', A, B, C and D instead of n, n', m, m', ℓ, ℓ', A, B, C, D, A', B', C' and D')). Thus, Lemma 7.207 (applied to n, m, n, m, I_n, $A'A + B'C$, $0_{n\times m}$, $A'B + B'D$, $0_{m\times n}$, $C'A + D'C$, I_m and $C'B + D'D$

instead of n, n', m, m', A, A', B, B', C, C', D and D') yields that $I_n = A'A + B'C$, $0_{n\times m} = A'B + B'D$, $0_{m\times n} = C'A + D'C$ and $I_m = C'B + D'D$.

Recall that $\det(I_n) = 1$. The same argument (applied to m instead of n) yields $\det(I_m) = 1$. But from $I_m = C'B + D'D$, we obtain $C'B + D'D = I_m$ and thus

$$\det \underbrace{\left(C'B + D'D\right)}_{=I_m} = \det(I_m) = 1.$$

From $0_{m\times n} = C'A + D'C$, we obtain $C'A = -D'C$. Hence, Exercise 6.35 (applied to $W = D'$ and $V = C'$) yields

$$\det\left(D'\right)\cdot\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \cdot \underbrace{\det\left(C'B + D'D\right)}_{=1} = \det A.$$

Thus,

$$\det A = \det\left(D'\right)\cdot\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det\begin{pmatrix} A & B \\ C & D \end{pmatrix}\cdot\det\left(D'\right).$$

This solves Exercise 6.37. □

7.106. Solution to Exercise 6.38

Our solution to Exercise 6.38 below will use the following simple lemma:

Lemma 7.208. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times m}$. Let $k \in \{0, 1, \ldots, n\}$ and $\ell \in \{0, 1, \ldots, m\}$. Then,

$$A = \begin{pmatrix} \operatorname{sub}^{1,2,\ldots,\ell}_{1,2,\ldots,k} A & \operatorname{sub}^{\ell+1,\ell+2,\ldots,m}_{1,2,\ldots,k} A \\ \operatorname{sub}^{1,2,\ldots,\ell}_{k+1,k+2,\ldots,n} A & \operatorname{sub}^{\ell+1,\ell+2,\ldots,m}_{k+1,k+2,\ldots,n} A \end{pmatrix}.$$

Proof of Lemma 7.208. In visual language, Lemma 7.208 states a triviality: It says that if we cut the matrix A horizontally (between its k-th and $(k+1)$-st rows) and vertically (between its ℓ-th and $(\ell+1)$-st columns), then we obtain four little matrices (namely, $\operatorname{sub}^{1,2,\ldots,\ell}_{1,2,\ldots,k} A$, $\operatorname{sub}^{\ell+1,\ell+2,\ldots,m}_{1,2,\ldots,k} A$, $\operatorname{sub}^{1,2,\ldots,\ell}_{k+1,k+2,\ldots,n} A$ and $\operatorname{sub}^{\ell+1,\ell+2,\ldots,m}_{k+1,k+2,\ldots,n} A$) which can be assembled back to form A (using the block-matrix construction). Turning this into a formal proof is straightforward. □

Solution to Exercise 6.38. Lemma 7.208 (applied to $m = n$ and $\ell = k$) yields

$$A = \begin{pmatrix} \operatorname{sub}^{1,2,\ldots,k}_{1,2,\ldots,k} A & \operatorname{sub}^{k+1,k+2,\ldots,n}_{1,2,\ldots,k} A \\ \operatorname{sub}^{1,2,\ldots,k}_{k+1,k+2,\ldots,n} A & \operatorname{sub}^{k+1,k+2,\ldots,n}_{k+1,k+2,\ldots,n} A \end{pmatrix}. \qquad (1276)$$

But the matrix $A \in \mathbb{K}^{n\times n}$ is invertible. Its inverse is $A^{-1} \in \mathbb{K}^{n\times n}$. Lemma 7.208 (applied to n, k and A^{-1} instead of m, ℓ and A) yields

$$A^{-1} = \begin{pmatrix} \operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A^{-1}\right) & \operatorname{sub}_{1,2,\ldots,k}^{k+1,k+2,\ldots,n}\left(A^{-1}\right) \\ \operatorname{sub}_{k+1,k+2,\ldots,n}^{1,2,\ldots,k}\left(A^{-1}\right) & \operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(A^{-1}\right) \end{pmatrix}. \tag{1277}$$

Now, recall that the matrix A is invertible, and its inverse is A^{-1}. In view of the equalities (1276) and (1277), this rewrites as follows: The matrix $\begin{pmatrix} \operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k} A & \operatorname{sub}_{1,2,\ldots,k}^{k+1,k+2,\ldots,n} A \\ \operatorname{sub}_{k+1,k+2,\ldots,n}^{1,2,\ldots,k} A & \operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n} A \end{pmatrix}$ is invertible, and its inverse is $\begin{pmatrix} \operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A^{-1}\right) & \operatorname{sub}_{1,2,\ldots,k}^{k+1,k+2,\ldots,n}\left(A^{-1}\right) \\ \operatorname{sub}_{k+1,k+2,\ldots,n}^{1,2,\ldots,k}\left(A^{-1}\right) & \operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(A^{-1}\right) \end{pmatrix}$. Hence, Exercise 6.37 (applied to k, $n-k$, $\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k} A$, $\operatorname{sub}_{1,2,\ldots,k}^{k+1,k+2,\ldots,n} A$, $\operatorname{sub}_{k+1,k+2,\ldots,n}^{1,2,\ldots,k} A$, $\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n} A$, $\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A^{-1}\right)$, $\operatorname{sub}_{1,2,\ldots,k}^{k+1,k+2,\ldots,n}\left(A^{-1}\right)$, $\operatorname{sub}_{k+1,k+2,\ldots,n}^{1,2,\ldots,k}\left(A^{-1}\right)$ and $\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(A^{-1}\right)$ instead of n, m, A, B, C, D, A', B', C' and D') yields

$$\begin{aligned} &\det\left(\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k} A\right) \\ &= \det \underbrace{\begin{pmatrix} \operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k} A & \operatorname{sub}_{1,2,\ldots,k}^{k+1,k+2,\ldots,n} A \\ \operatorname{sub}_{k+1,k+2,\ldots,n}^{1,2,\ldots,k} A & \operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n} A \end{pmatrix}}_{\substack{=A \\ \text{(by (1276))}}} \cdot \det\left(\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(A^{-1}\right)\right) \\ &= \det A \cdot \det\left(\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(A^{-1}\right)\right). \end{aligned}$$

This solves Exercise 6.38. □

7.107. Solution to Exercise 6.39

Solution to Exercise 6.39. **(a)** All claims of Proposition 6.130 and Proposition 6.131 are "clear by inspection", in the sense that the reader will have no difficulty convincing themselves of their validity just by studying an example and seeing "what is going on". Let me give some more detailed proofs. (The following proofs are not very formal; but I will outline a formal proof of Proposition 6.130 **(h)** in a footnote, and I trust that the reader can use the same methods to formally prove the rest of Proposition 6.130 if so inclined.)

Proof of Proposition 6.130. Proposition 6.130 **(a)** follows from the definitions of $A_{u,\bullet}$ and of $\operatorname{rows}_u A$ (indeed, these definitions show that both $A_{u,\bullet}$ and $\operatorname{rows}_u A$ are the u-th row of A). Similarly, Proposition 6.130 **(b)** follows from the definitions of $A_{\bullet,v}$ and $\operatorname{cols}_v A$.

Proposition 6.130 **(c)** is obvious[562].

Proposition 6.130 **(d)** claims that the w-th column of the matrix $A_{\bullet,\sim v}$ equals the w-th column of the matrix A (whenever $v \in \{1, 2, \ldots, m\}$ and $w \in \{1, 2, \ldots, v-1\}$). Let us prove this: Let $v \in \{1, 2, \ldots, m\}$ and $w \in \{1, 2, \ldots, v-1\}$. Thus, $w < v$ (since $w \in \{1, 2, \ldots, v-1\}$). But the matrix $A_{\bullet,\sim v}$ results from the matrix A by removing the v-th column; clearly, this removal does not change the w-th column (because $w < v$). Thus, the w-th column of the matrix $A_{\bullet,\sim v}$ equals the w-th column of the

[562] In fact:

- The matrix $(A_{\bullet,\sim v})_{\sim u,\bullet}$ is the matrix obtained from A by first removing the v-th column and then removing the u-th row.
- The matrix $(A_{\sim u,\bullet})_{\bullet,\sim v}$ is the matrix obtained from A by first removing the u-th row and then removing the v-th column.
- The matrix $A_{\sim u,\sim v}$ is the matrix obtained from A by removing the u-th row and the v-th column at the same time.

Thus it is clear that all three of these matrices are equal.

matrix A. So Proposition 6.130 **(d)** is proven.[563]

Proposition 6.130 **(e)** claims that the w-th column of the matrix $A_{\bullet,\sim v}$ equals

[563]If you found this proof insufficiently rigorous, let me show a formal proof of Proposition 6.130 **(d)**. First, I shall introduce a notation:

- For every $j \in \mathbb{Z}$ and $r \in \mathbb{Z}$, let $\mathbf{d}_r(j)$ be the integer $\begin{cases} j, & \text{if } j < r; \\ j+1, & \text{if } j \geq r \end{cases}$.

We make the following observation:

Observation 1: If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, if $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} \in \mathbb{K}^{n\times m}$ is an $n \times m$-matrix, and if v is an element of $\{1,2,\ldots,m\}$, then $A_{\bullet,\sim v} = \left(a_{i,\mathbf{d}_v(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq m-1}$.

Proof of Observation 1: Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} \in \mathbb{K}^{n\times m}$ be an $n \times m$-matrix. Let v be an element of $\{1,2,\ldots,m\}$. Recalling how $\mathbf{d}_v(j)$ is defined for every $j \in \mathbb{Z}$, we see that

$$\begin{aligned}(\mathbf{d}_v(1), \mathbf{d}_v(2), \ldots, \mathbf{d}_v(m-1)) &= (1,2,\ldots,v-1,v+1,v+2,\ldots,m) \\ &= (1,2,\ldots,\widehat{v},\ldots,m).\end{aligned}$$

But the definition of $A_{\bullet,\sim v}$ yields

$$\begin{aligned}A_{\bullet,\sim v} &= \operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,m} A = \operatorname{cols}_{\mathbf{d}_v(1),\mathbf{d}_v(2),\ldots,\mathbf{d}_v(m-1)} A \\ &\qquad \left(\text{since } (1,2,\ldots,\widehat{v},\ldots,m) = (\mathbf{d}_v(1), \mathbf{d}_v(2), \ldots, \mathbf{d}_v(m-1))\right) \\ &= \left(a_{i,\mathbf{d}_v(y)}\right)_{1\leq i\leq n,\ 1\leq y\leq m-1} \qquad \left(\begin{array}{c}\text{by the definition of } \operatorname{cols}_{\mathbf{d}_v(1),\mathbf{d}_v(2),\ldots,\mathbf{d}_v(m-1)} A \\ \text{(since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}\text{)}\end{array}\right) \\ &= \left(a_{i,\mathbf{d}_v(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq m-1} \qquad \left(\text{here, we have renamed the index } (i,y) \text{ as } (i,j)\right).\end{aligned}$$

This proves Observation 1.

Let us now prove Proposition 6.130 **(d)**. Indeed, let $v \in \{1,2,\ldots,m\}$ and $w \in \{1,2,\ldots,v-1\}$. Then, $w < v$ (since $w \in \{1,2,\ldots,v-1\}$). Write the $n \times m$-matrix A as $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$. The definition of $\mathbf{d}_v(w)$ now yields $\mathbf{d}_v(w) = \begin{cases} w, & \text{if } w < v; \\ w+1, & \text{if } w \geq v \end{cases} = w$ (since $w < v$). Now, the definition of $\left(A_{\bullet,\sim v}\right)_{\bullet,w}$ yields

$$\begin{aligned}\left(A_{\bullet,\sim v}\right)_{\bullet,w} &= \left(\text{the } w\text{-th column of the matrix } A_{\bullet,\sim v}\right) \\ &= \begin{pmatrix} a_{1,\mathbf{d}_v(w)} \\ a_{2,\mathbf{d}_v(w)} \\ \vdots \\ a_{n,\mathbf{d}_v(w)} \end{pmatrix} \qquad \left(\text{since Observation 1 yields } A_{\bullet,\sim v} = \left(a_{i,\mathbf{d}_v(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq m-1}\right) \\ &= \begin{pmatrix} a_{1,w} \\ a_{2,w} \\ \vdots \\ a_{n,w} \end{pmatrix} \qquad \left(\text{since } \mathbf{d}_v(w) = w\right) \\ &= \left(\text{the } w\text{-th column of the matrix } A\right) \qquad \left(\text{since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}\right) \\ &= A_{\bullet,w}.\end{aligned}$$

the $(w+1)$-th column of the matrix A (whenever $v \in \{1,2,\ldots,m\}$ and $w \in \{v,v+1,\ldots,m-1\}$). Let us prove this:

Let $v \in \{1,2,\ldots,m\}$ and $w \in \{v,v+1,\ldots,m-1\}$. Thus, $w+1 > w \geq v$ (since $w \in \{v,v+1,\ldots,m-1\}$). But the matrix $A_{\bullet,\sim v}$ results from the matrix A by removing the v-th column; this removal moves the $(w+1)$-th column of A one step leftwards (since $w+1 > v$). In other words, the $(w+1)$-th column of A becomes the w-th column of the new matrix $A_{\bullet,\sim v}$. In other words, the w-th column of the matrix $A_{\bullet,\sim v}$ equals the $(w+1)$-th column of the matrix A. So Proposition 6.130 **(e)** is proven.[564]

Proposition 6.130 **(f)** is the analogue of Proposition 6.130 **(d)** for rows instead of columns (with v renamed as u); its proof is equally analogous.

Proposition 6.130 **(g)** is the analogue of Proposition 6.130 **(e)** for rows instead of columns (with v renamed as u); its proof is equally analogous.

Let us now prove Proposition 6.130 **(h)**: Let $v \in \{1,2,\ldots,m\}$ and $w \in \{1,2,\ldots,v-1\}$. We need to show that $(A_{\bullet,\sim v})_{\bullet,\sim w} = \operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A$. The matrix $(A_{\bullet,\sim v})_{\bullet,\sim w}$ is obtained from the matrix A by first removing the v-th column (thus obtaining an $n \times (m-1)$-matrix) and then removing the w-th column (from the resulting $n \times (m-1)$-matrix). Since $w < v$ (because $w \in \{1,2,\ldots,v-1\}$), the removal of the v-th column of A did not affect the w-th column, and therefore the two successive removals could be replaced by a simultaneous removal of both the w-th and the v-th columns from the matrix A. In other words, the matrix obtained from the matrix A by first removing the v-th column and then removing the w-th column is identical with the matrix obtained from A by simultaneously removing both the w-th and the v-th columns. But the former matrix is $(A_{\bullet,\sim v})_{\bullet,\sim w}$, whereas the latter matrix is $\operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A$. Hence, the identity of these two matrices rewrites as $(A_{\bullet,\sim v})_{\bullet,\sim w} = \operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A$. This proves Proposition 6.130 **(h)**.[565]

This proves Proposition 6.130 **(d)**.

[564] Again, a more rigorous proof of Proposition 6.130 **(e)** can be obtained similarly to our rigorous proof of Proposition 6.130 **(d)** in the previous footnote. (The main difference is that we have $\mathbf{d}_v(w) = w+1$ instead of $\mathbf{d}_v(w) = w$ this time.)

[565] If you found this proof insufficiently rigorous, let me show a formal proof of Proposition 6.130 **(h)**. First, we have $r \leq s-1$ (since $r < s$ and since r and s are integers). Now, I shall introduce two notations:

- For every $j \in \mathbb{Z}$ and $r \in \mathbb{Z}$, let $\mathbf{d}_r(j)$ be the integer $\begin{cases} j, & \text{if } j < r; \\ j+1, & \text{if } j \geq r \end{cases}$.
- For every $j \in \mathbb{Z}$, $r \in \mathbb{Z}$ and $s \in \mathbb{Z}$ satisfying $r < s$, we let $\mathbf{d}_{r,s}(j)$ be the integer $\begin{cases} j, & \text{if } j < r; \\ j+1, & \text{if } r \leq j < s-1; \\ j+2, & \text{if } s-1 \leq j \end{cases}$. (This is well-defined, because $r \leq s-1$.)

We make the following three observations:

Observation 1: If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, if $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m} \in \mathbb{K}^{n \times m}$ is an $n \times m$-matrix, and if v is an element of $\{1,2,\ldots,m\}$, then $A_{\bullet,\sim v} = \left(a_{i,\mathbf{d}_v(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m-1}$.

Observation 2: If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, if $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m} \in \mathbb{K}^{n \times m}$ is an $n \times m$-matrix, and if w and v are two elements of $\{1, 2, \ldots, m\}$ satisfying $w < v$, then $\operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A = \left(a_{i,\mathbf{d}_{w,v}(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m-2}$.

Observation 3: For every two integers w and v satisfying $w < v$, we have

$$\mathbf{d}_{w,v}(j) = \mathbf{d}_v(\mathbf{d}_w(j)) \qquad \text{for each } j \in \mathbb{Z}.$$

Proof of Observation 1: Observation 1 was already proven in a previous footnote (namely, the one where we gave a rigorous proof of Proposition 6.130 **(d)**).

Proof of Observation 2: Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m} \in \mathbb{K}^{n \times m}$ be an $n \times m$-matrix. Let w and v be two elements of $\{1, 2, \ldots, m\}$ satisfying $w < v$. Recalling how $\mathbf{d}_{w,v}(j)$ is defined for every $j \in \mathbb{Z}$, we find that

$$\begin{aligned}
&(\mathbf{d}_{w,v}(1), \mathbf{d}_{w,v}(2), \ldots, \mathbf{d}_{w,v}(m-2)) \\
&= (1, 2, \ldots, w-1, w+1, w+2, \ldots, v-1, v+1, v+2, \ldots, m) \\
&= (1, 2, \ldots, \widehat{w}, \ldots, \widehat{v}, \ldots, m).
\end{aligned}$$

Now,

$$\begin{aligned}
&\operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A \\
&= \operatorname{cols}_{\mathbf{d}_{w,v}(1),\mathbf{d}_{w,v}(2),\ldots,\mathbf{d}_{w,v}(m-2)} A \\
&\qquad \left(\text{since } (1, 2, \ldots, \widehat{w}, \ldots, \widehat{v}, \ldots, m) = (\mathbf{d}_{w,v}(1), \mathbf{d}_{w,v}(2), \ldots, \mathbf{d}_{w,v}(m-2))\right) \\
&= \left(a_{i,\mathbf{d}_{w,v}(y)}\right)_{1 \leq i \leq n,\ 1 \leq y \leq m-2} \\
&\qquad \left(\begin{array}{c}\text{by the definition of } \operatorname{cols}_{\mathbf{d}_{w,v}(1),\mathbf{d}_{w,v}(2),\ldots,\mathbf{d}_{w,v}(m-2)} A \\ (\text{since } A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m})\end{array}\right) \\
&= \left(a_{i,\mathbf{d}_{w,v}(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m-2} \qquad \left(\text{here, we have renamed the index } (i, y) \text{ as } (i, j)\right).
\end{aligned}$$

This proves Observation 2.

Proof of Observation 3: Let w and v be two integers satisfying $w < v$. Then, $w \leq v - 1$ (since w and v are integers). For every $j \in \mathbb{Z}$, we can prove $\mathbf{d}_{w,v}(j) = \mathbf{d}_v(\mathbf{d}_w(j))$ by an explicit computation using the definitions of $\mathbf{d}_{w,v}(j)$, of $\mathbf{d}_w(j)$ and of $\mathbf{d}_v(\mathbf{d}_w(j))$. (The cases when $j < w$, when $w \leq j < v - 1$ and when $v - 1 \leq j$ need to be treated separately; but each of these three cases is completely straightforward. For instance, in the case when $v - 1 \leq j$, the definition of $\mathbf{d}_{w,v}(j)$ yields

$$\mathbf{d}_{w,v}(j) = \begin{cases} j, & \text{if } j < w; \\ j+1, & \text{if } w \leq j < v-1; \\ j+2, & \text{if } v-1 \leq j \end{cases} = j + 2 \qquad (\text{since } v - 1 \leq j),$$

whereas the definition of $\mathbf{d}_w(j)$ yields

$$\mathbf{d}_w(j) = \begin{cases} j, & \text{if } j < w; \\ j+1, & \text{if } j \geq w \end{cases} = j + 1 \qquad (\text{since } j \geq v - 1 \geq w)$$

Similarly, we can prove Proposition 6.130 **(i)**[566]: Let $v \in \{1,2,\ldots,m\}$ and $w \in \{v,v+1,\ldots,m-1\}$. We need to show that $\left(A_{\bullet,\sim v}\right)_{\bullet,\sim w} = \operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w+1},\ldots,m} A$. The matrix $\left(A_{\bullet,\sim v}\right)_{\bullet,\sim w}$ is obtained from the matrix A by first removing the v-th column (thus obtaining an $n \times (m-1)$-matrix) and then removing the w-th column (from the resulting $n \times (m-1)$-matrix). Since $w \geq v$ (because $w \in \{v,v+1,\ldots,m-1\}$), the first of these two removals (i.e., the removal of the v-th column of A) has moved the $(w+1)$-th column of A one step to the left; therefore, the latter column has become the w-th column after this first removal. The second removal then removes this column. Therefore, the two successive removals could be replaced by a simultaneous removal of both the v-th and the $(w+1)$-th columns from the matrix A. In other words, the matrix obtained from the matrix A by first removing the v-th column and then removing the w-th column is identical with the matrix obtained from A by simultaneously removing both the v-th and the $(w+1)$-th columns. But the former matrix is $\left(A_{\bullet,\sim v}\right)_{\bullet,\sim w}$, whereas the latter matrix is $\operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w+1},\ldots,m} A$. Hence, the identity of these two matrices rewrites as $\left(A_{\bullet,\sim v}\right)_{\bullet,\sim w} = \operatorname{cols}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w+1},\ldots,m} A$. This proves Proposition 6.130 **(i)**.[567]

---

and therefore

$$\begin{aligned} \mathbf{d}_v\left(\mathbf{d}_w\left(j\right)\right) &= \mathbf{d}_v\left(j+1\right) = \begin{cases} j+1, & \text{if } j+1 < v; \\ (j+1)+1, & \text{if } j+1 \geq v \end{cases} \qquad \left(\text{by the definition of } \mathbf{d}_v\left(j+1\right)\right) \\ &= (j+1)+1 \qquad \left(\text{since } j+1 \geq v \text{ (since } j \geq v-1\text{)}\right) \\ &= j+2 = \mathbf{d}_{w,v}\left(j\right); \end{aligned}$$

thus, $\mathbf{d}_{w,v}\left(j\right) = \mathbf{d}_v\left(\mathbf{d}_w\left(j\right)\right)$ is proven in this case.) This proves Observation 3.

Let us now prove Proposition 6.130 **(h)**. Indeed, let $v \in \{1,2,\ldots,m\}$ and $w \in \{1,2,\ldots,v-1\}$. Then, $w \leq v-1$ (since $w \in \{1,2,\ldots,v-1\}$) and thus $w \leq \underbrace{v}_{\leq m} -1 \leq m-1$, so that $v \in \{1,2,\ldots,m-1\}$.

Write the matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$. Then, Observation 1 shows that $A_{\bullet,\sim v} = \left(a_{i,\mathbf{d}_v(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq m-1}$. Hence, Observation 1 (applied to $m-1$, $A_{\bullet,\sim v}$, $a_{i,\mathbf{d}_v(j)}$ and w instead of m, A, $a_{i,j}$ and v) yields

$$\begin{aligned} \left(A_{\bullet,\sim v}\right)_{\bullet,\sim w} &= \left(a_{i,\mathbf{d}_v(\mathbf{d}_w(j))}\right)_{1\leq i\leq n,\ 1\leq j\leq (m-1)-1} = \left(\underbrace{a_{i,\mathbf{d}_v(\mathbf{d}_w(j))}}_{\substack{=a_{i,\mathbf{d}_{w,v}(j)} \\ \text{(since Observation 3 yields} \\ \mathbf{d}_v(\mathbf{d}_w(j))=\mathbf{d}_{w,v}(j))}}\right)_{1\leq i\leq n,\ 1\leq j\leq m-2} \\ &= \left(a_{i,\mathbf{d}_{w,v}(j)}\right)_{1\leq i\leq n,\ 1\leq j\leq m-2} = \operatorname{cols}_{1,2,\ldots,\widehat{w},\ldots,\widehat{v},\ldots,m} A \end{aligned}$$

(by Observation 2). This proves Proposition 6.130 **(h)**.

[566] We will not actually need Proposition 6.130 **(i)** in the following, so you can just as well skip this proof. The same applies to Proposition 6.130 **(k)**.

[567] Again, a more rigorous proof can be given (if desired) similarly to our rigorous proof of Proposi-

Proposition 6.130 **(j)** is the analogue of Proposition 6.130 **(h)** for rows instead of columns (with v renamed as u); its proof is equally analogous.

Proposition 6.130 **(k)** is the analogue of Proposition 6.130 **(i)** for rows instead of columns (with v renamed as u); its proof is equally analogous.

Let us now prove Proposition 6.130 **(l)**: Let $v \in \{1, 2, \ldots, n\}$, $u \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, m\}$ be such that $u < v$. We have $u < v$ and thus $u \leq v - 1$ (since u and v are integers). Combined with $u \geq 1$ (since $u \in \{1, 2, \ldots, n\}$), this yields $u \in \{1, 2, \ldots, v - 1\}$.

Proposition 6.130 **(c)** (applied to v and q instead of u and v) yields $\left(A_{\bullet,\sim q}\right)_{\sim v,\bullet} = \left(A_{\sim v,\bullet}\right)_{\bullet,\sim q} = A_{\sim v,\sim q}$.

Now, $u \leq \underbrace{v}_{\substack{\leq n \\ (\text{since } v \in \{1,2,\ldots,n\})}} -1 \leq n - 1$. Combining this with $u \geq 1$, we obtain $u \in \{1, 2, \ldots, n - 1\}$. Hence, Proposition 6.130 **(c)** (applied to $n - 1$, $A_{\sim v,\bullet}$ and q instead of n, A and v) yields $\left(\left(A_{\sim v,\bullet}\right)_{\bullet,\sim q}\right)_{\sim u,\bullet} = \left(\left(A_{\sim v,\bullet}\right)_{\sim u,\bullet}\right)_{\bullet,\sim q} = \left(A_{\sim v,\bullet}\right)_{\sim u,\sim q}$. Hence,

$$\left(A_{\sim v,\bullet}\right)_{\sim u,\sim q} = \left(\underbrace{\left(A_{\sim v,\bullet}\right)_{\bullet,\sim q}}_{=\left(A_{\bullet,\sim q}\right)_{\sim v,\bullet}}\right)_{\sim u,\bullet} = \left(\left(A_{\bullet,\sim q}\right)_{\sim v,\bullet}\right)_{\sim u,\bullet}$$
$$= \operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(A_{\bullet,\sim q}\right)$$

(by Proposition 6.130 **(j)**, applied to $m - 1$, $A_{\bullet,\sim q}$, v and u instead of m, A, u and w). This proves Proposition 6.130 **(l)**. $\square$

Thus, the proof of Proposition 6.130 is complete. It remains to prove Proposition 6.131. Let me give an informal proof:

Proof of Proposition 6.131. We know that $\left(I_n\right)_{\bullet,u}$ is the u-th column of the identity matrix. All entries of this column are 0, except for the u-th, which is 1. In other words,

$$\left(I_n\right)_{\bullet,u} = \left(\underbrace{0, 0, \ldots, 0}_{u-1 \text{ zeroes}}, 1, \underbrace{0, 0, \ldots, 0}_{n-u \text{ zeroes}}\right)^T .$$

Now, $\left(\left(I_n\right)_{\bullet,u}\right)_{\sim v,\bullet}$ is the result of removing the v-th row from this column, i.e., removing the v-th entry from this column[568]. This v-th row is below the 1 in the

tion 6.130 **(h)**.

[568] Indeed, the v-th row is the same as the v-th entry in this case, because each row of $\left(I_n\right)_{\bullet,u}$ consists of one entry only.

u-th row (since $u < v$); therefore, the result of its removal is the column vector

$$\left(\underbrace{0,0,\ldots,0}_{u-1\text{ zeroes}},1,\underbrace{0,0,\ldots,0}_{n-u-1\text{ zeroes}}\right)^T.$$

Hence, we have shown that

$$\begin{aligned}\left((I_n)_{\bullet,u}\right)_{\sim v,\bullet} &= \left(\underbrace{0,0,\ldots,0}_{u-1\text{ zeroes}},1,\underbrace{0,0,\ldots,0}_{n-u-1\text{ zeroes}}\right)^T \\ &= \left(\underbrace{0,0,\ldots,0}_{u-1\text{ zeroes}},1,\underbrace{0,0,\ldots,0}_{(n-1)-u\text{ zeroes}}\right)^T \\ &= \left(\text{the } u\text{-th column of the matrix } I_{n-1}\right) = (I_{n-1})_{\bullet,u}.\end{aligned}$$

This proves Proposition 6.131[569]. □

[569] Again, if this proof was not rigorous enough for you, here is a more formal proof of Proposition 6.131:

For every $j \in \mathbb{Z}$ and $r \in \mathbb{Z}$, let $\mathbf{d}_r(j)$ be the integer $\begin{cases} j, & \text{if } j < r; \\ j+1, & \text{if } j \geq r \end{cases}$. For any two objects i and j, we define an element $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

We make the following two observations:

Observation 1: If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, if $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m} \in \mathbb{K}^{n\times m}$ is an $n \times m$-matrix, and if u is an element of $\{1, 2, \ldots, n\}$, then $A_{\sim u,\bullet} = \left(a_{\mathbf{d}_u(i),j}\right)_{1\leq i\leq n-1,\ 1\leq j\leq m}$.

Observation 2: If $u \in \mathbb{Z}$, $v \in \mathbb{Z}$ and $i \in \mathbb{Z}$ satisfy $u < v$, then $\delta_{\mathbf{d}_v(i),u} = \delta_{i,u}$.

Proof of Observation 1: Observation 1 is analogous to the Observation 1 from the footnote in our above proof of Proposition 6.130 **(h)**. (More precisely, it is the analogue of the latter observation for rows instead of columns.)

Proof of Observation 2: Let $u \in \mathbb{Z}$, $v \in \mathbb{Z}$ and $i \in \mathbb{Z}$ satisfy $u < v$. We are in one of the following two Cases:

Case 1: We have $i < v$.

Case 2: We have $i \geq v$.

Let us first consider Case 1. In this case, we have $i < v$. Now, the definition of $\mathbf{d}_v(i)$ yields $\mathbf{d}_v(i) = \begin{cases} i, & \text{if } i < v; \\ i+1, & \text{if } i \geq v \end{cases} = i$ (since $i < v$) and thus $\delta_{\mathbf{d}_v(i),u} = \delta_{i,u}$. Hence, Observation 2 is proven in Case 1.

Let us now consider Case 2. In this case, we have $i \geq v$. Now, the definition of $\mathbf{d}_v(i)$ yields $\mathbf{d}_v(i) = \begin{cases} i, & \text{if } i < v; \\ i+1, & \text{if } i \geq v \end{cases} = i+1$ (since $i \geq v$) and thus $\mathbf{d}_v(i) = i+1 > i \geq v > u$. Hence, $\mathbf{d}_v(i) \neq u$. Thus, $\delta_{\mathbf{d}_v(i),u} = 0$. Also, $i \geq v > u$ and thus $i \neq u$; hence, $\delta_{i,u} = 0$. Now, $\delta_{\mathbf{d}_v(i),u} = 0 = \delta_{i,u}$. Hence, Observation 2 is proven in Case 2.

Thus, Observation 2 is proven in both Cases 1 and 2; hence, Observation 2 is always proven.

Now, let n, u and v be as in Proposition 6.131. From $u < v \leq n$, we obtain $u \in \{1, 2, \ldots, n-1\}$, We have now proven both Proposition 6.130 and Proposition 6.131. Thus, Exercise 6.39 **(a)** is solved.

(b) We need to derive Proposition 6.123 and Proposition 6.122 from Theorem 6.126.

Proof of Proposition 6.123 using Theorem 6.126. We have

$$\operatorname{sub}_{1,2,\ldots,\widehat{1},\ldots,\widehat{2},\ldots,n}^{1,2,\ldots,\widehat{1},\ldots,\widehat{2},\ldots,n} A = \widetilde{A} \tag{1278}$$

[570]. Now, $1 < 2$. Moreover, 1 and 2 are elements of $\{1, 2, \ldots, n\}$ (since $n \geq 2$).

---

so that $(I_{n-1})_{\bullet,u}$ is well-defined. Also, $\left((I_n)_{\bullet,u}\right)_{\sim v,\bullet}$ is well-defined (since u and v belong to $\{1, 2, \ldots, n\}$). Now, the definition of $(I_n)_{\bullet,u}$ yields

$$\begin{aligned} (I_n)_{\bullet,u} &= (\text{the } u\text{-th column of the matrix } I_n) \\ &= \begin{pmatrix} \delta_{1,u} \\ \delta_{2,u} \\ \vdots \\ \delta_{n,u} \end{pmatrix} \qquad \left(\text{since } I_n = \left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ &= \left(\delta_{i,u}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}. \end{aligned}$$

The same argument (applied to $n-1$ instead of n) yields $(I_{n-1})_{\bullet,u} = \left(\delta_{i,u}\right)_{1\leq i\leq n-1,\ 1\leq j\leq 1}$.

Recall that $(I_n)_{\bullet,u} = \left(\delta_{i,u}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}$. Hence, Observation 1 (applied to 1, $(I_n)_{\bullet,u}$, $\delta_{i,u}$ and v instead of m, A, $a_{i,j}$ and u) yields

$$\left((I_n)_{\bullet,u}\right)_{\sim v,\bullet} = \left(\underbrace{\delta_{\mathbf{d}_v(i),u}}_{\substack{=\delta_{i,u} \\ \text{(by Observation 2)}}} \right)_{1\leq i\leq n-1,\ 1\leq j\leq 1} = \left(\delta_{i,u}\right)_{1\leq i\leq n-1,\ 1\leq j\leq 1}.$$

Comparing this with $(I_{n-1})_{\bullet,u} = \left(\delta_{i,u}\right)_{1\leq i\leq n-1,\ 1\leq j\leq 1}$, we obtain $\left((I_n)_{\bullet,u}\right)_{\sim v,\bullet} = (I_{n-1})_{\bullet,u}$. This proves Proposition 6.131.

[570]*Proof of (1278):* The $(n-2)$-tuple $\left(1, 2, \ldots, \widehat{1}, \ldots, \widehat{2}, \ldots, n\right)$ is obtained by removing the 1-st and the 2-nd entries from the n-tuple $(1, 2, \ldots, n)$. Thus,

$$\left(1, 2, \ldots, \widehat{1}, \ldots, \widehat{2}, \ldots, n\right) = (3, 4, \ldots, n) = (1+2, 2+2, \ldots, (n-2)+2).$$

Hence,

$$\begin{aligned} \operatorname{sub}_{1,2,\ldots,\widehat{1},\ldots,\widehat{2},\ldots,n}^{1,2,\ldots,\widehat{1},\ldots,\widehat{2},\ldots,n} A &= \operatorname{sub}_{1+2,2+2,\ldots,(n-2)+2}^{1+2,2+2,\ldots,(n-2)+2} A = \left(a_{x+2,y+2}\right)_{1\leq x\leq n-2,\ 1\leq y\leq n-2} \\ &\qquad \left(\text{by the definition of } \operatorname{sub}_{1+2,2+2,\ldots,(n-2)+2}^{1+2,2+2,\ldots,(n-2)+2} A\right) \\ &= \left(a_{i+2,j+2}\right)_{1\leq i\leq n-2,\ 1\leq j\leq n-2} \\ &\qquad (\text{here, we have renamed the index } (x,y) \text{ as } (i,j)) \\ &= \widetilde{A} \qquad \left(\text{since } \widetilde{A} = \left(a_{i+2,j+2}\right)_{1\leq i\leq n-2,\ 1\leq j\leq n-2}\right). \end{aligned}$$

Hence, Theorem 6.126 (applied to $p = 1$, $q = 2$, $u = 1$ and $v = 2$) yields

$$\det A \cdot \det \left(\operatorname{sub}^{1,2,\ldots,\widehat{1},\ldots,\widehat{2},\ldots,n}_{1,2,\ldots,\widehat{1},\ldots,\widehat{2},\ldots,n} A \right) = \det \left(A_{\sim 1,\sim 1} \right) \cdot \det \left(A_{\sim 2,\sim 2} \right) - \det \left(A_{\sim 1,\sim 2} \right) \cdot \det \left(A_{\sim 2,\sim 1} \right).$$

In view of (1278), this rewrites as

$$\det A \cdot \det \widetilde{A} = \det \left(A_{\sim 1,\sim 1} \right) \cdot \det \left(A_{\sim 2,\sim 2} \right) - \det \left(A_{\sim 1,\sim 2} \right) \cdot \det \left(A_{\sim 2,\sim 1} \right).$$

This proves Proposition 6.123. □

Proof of Proposition 6.122 using Theorem 6.126. We have

$$\operatorname{sub}^{1,2,\ldots,\widehat{1},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{1},\ldots,\widehat{n},\ldots,n} A = A' \qquad (1279)$$

[571]. Now, $1 < n$ (since $n \geq 2$) and $1 < n$. Moreover, 1 and n are elements of $\{1, 2, \ldots, n\}$ (since $n \geq 2 \geq 1$). Hence, Theorem 6.126 (applied to $p = 1$, $q = n$, $u = 1$ and $v = n$) yields

$$\det A \cdot \det \left(\operatorname{sub}^{1,2,\ldots,\widehat{1},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{1},\ldots,\widehat{n},\ldots,n} A \right) = \det \left(A_{\sim 1,\sim 1} \right) \cdot \det \left(A_{\sim n,\sim n} \right) - \det \left(A_{\sim 1,\sim n} \right) \cdot \det \left(A_{\sim n,\sim 1} \right).$$

In view of (1279), this rewrites as

$$\det A \cdot \det \left(A' \right) = \det \left(A_{\sim 1,\sim 1} \right) \cdot \det \left(A_{\sim n,\sim n} \right) - \det \left(A_{\sim 1,\sim n} \right) \cdot \det \left(A_{\sim n,\sim 1} \right).$$

This proves Proposition 6.122. □

Now, both Proposition 6.123 and Proposition 6.122 are proven using Theorem 6.126. This completes the solution of Exercise 6.39 **(b)**. □

Qed.

[571] *Proof of (1279):* The $(n-2)$-tuple $\left(1, 2, \ldots, \widehat{1}, \ldots, \widehat{n}, \ldots, n\right)$ is obtained by removing the 1-st and the n-th entries from the n-tuple $(1, 2, \ldots, n)$. Thus,

$$\left(1, 2, \ldots, \widehat{1}, \ldots, \widehat{n}, \ldots, n\right) = (2, 3, \ldots, n-1) = (1+1, 2+1, \ldots, (n-2)+1).$$

Hence,

$$\begin{aligned} \operatorname{sub}^{1,2,\ldots,\widehat{1},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{1},\ldots,\widehat{n},\ldots,n} A = \operatorname{sub}^{1+1,2+1,\ldots,(n-2)+1}_{1+1,2+1,\ldots,(n-2)+1} A &= \left(a_{x+1,y+1}\right)_{1 \leq x \leq n-2,\ 1 \leq y \leq n-2} \\ &\qquad \left(\text{by the definition of } \operatorname{sub}^{1+1,2+1,\ldots,(n-2)+1}_{1+1,2+1,\ldots,(n-2)+1} A\right) \\ &= \left(a_{i+1,j+1}\right)_{1 \leq i \leq n-2,\ 1 \leq j \leq n-2} \\ &\qquad \left(\text{here, we have renamed the index } (x,y) \text{ as } (i,j)\right) \\ &= A' \qquad \left(\text{since } A' = \left(a_{i+1,j+1}\right)_{1 \leq i \leq n-2,\ 1 \leq j \leq n-2}\right). \end{aligned}$$

Qed.

7.108. Solution to Exercise 6.40

Proof of Proposition 6.134. The matrix $(A \mid v)$ is defined as the $n \times (m+1)$-matrix whose $m+1$ columns are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}, v$ (from left to right). Thus, the first m columns of the matrix $(A \mid v)$ are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}$, whereas the $(m+1)$-st column of $(A \mid v)$ is v.

(a) For every $q \in \{1, 2, \ldots, m\}$, we have

$$\begin{aligned} (A \mid v)_{\bullet,q} &= (\text{the } q\text{-th column of the matrix } (A \mid v)) \\ &\qquad \left(\text{by the definition of } (A \mid v)_{\bullet,q}\right) \\ &= A_{\bullet,q} \end{aligned}$$

(since the first m columns of the matrix $(A \mid v)$ are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}$). This proves Proposition 6.134 **(a)**.

(b) The definition of $(A \mid v)_{\bullet,m+1}$ yields

$$(A \mid v)_{\bullet,m+1} = (\text{the } (m+1)\text{-st column of the matrix } (A \mid v)) = v$$

(since the $(m+1)$-st column of $(A \mid v)$ is v). This proves Proposition 6.134 **(b)**.

(c) Let $q \in \{1, 2, \ldots, m\}$. Recall that $(A \mid v)$ is the $n \times (m+1)$-matrix whose $m+1$ columns are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}, v$. The matrix $(A \mid v)_{\bullet,\sim q}$ is obtained from this matrix $(A \mid v)$ by removing its q-th column; thus,

$$\text{the columns of this matrix } (A \mid v)_{\bullet,\sim q} \text{ are } A_{\bullet,1}, A_{\bullet,2}, \ldots, \widehat{A_{\bullet,q}}, \ldots, A_{\bullet,m}, v \qquad (1280)$$

(since the columns of the matrix $(A \mid v)$ are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}, v$, and since $q \in \{1, 2, \ldots, m\}$).

On the other hand, the matrix $A_{\bullet,\sim q}$ is obtained from the matrix A by removing its q-th column; thus, the columns of this matrix $A_{\bullet,\sim q}$ are $A_{\bullet,1}, A_{\bullet,2}, \ldots, \widehat{A_{\bullet,q}}, \ldots, A_{\bullet,m}$. Hence,

$$\text{the columns of the matrix } \left(A_{\bullet,\sim q} \mid v\right) \text{ are } A_{\bullet,1}, A_{\bullet,2}, \ldots, \widehat{A_{\bullet,q}}, \ldots, A_{\bullet,m}, v \qquad (1281)$$

(since the matrix $\left(A_{\bullet,\sim q} \mid v\right)$ is obtained from $A_{\bullet,\sim q}$ by attaching the column v at its "right edge").

Comparing (1280) with (1281), we see that the columns of the matrix $(A \mid v)_{\bullet,\sim q}$ are precisely the columns of the matrix $\left(A_{\bullet,\sim q} \mid v\right)$. Thus, these two matrices must be identical. In other words, $(A \mid v)_{\bullet,\sim q} = \left(A_{\bullet,\sim q} \mid v\right)$. This proves Proposition 6.134 **(c)**.

(d) We have defined $(A \mid v)$ as the $n \times (m+1)$-matrix whose $m+1$ columns are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}, v$. The matrix $(A \mid v)_{\bullet,\sim(m+1)}$ is obtained from this matrix $(A \mid v)$ by removing its $(m+1)$-th column; thus,

$$\text{the columns of this matrix } (A \mid v)_{\bullet,\sim(m+1)} \text{ are } A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m} \qquad (1282)$$

(since the columns of the matrix $(A \mid v)$ are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}, v$). On the other hand, clearly,

$$\text{the columns of the matrix } A \text{ are } A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}. \qquad (1283)$$

Comparing (1282) with (1283), we see that the columns of the matrix $(A \mid v)_{\bullet,\sim(m+1)}$ are precisely the columns of the matrix A. Thus, these two matrices must be identical. In other words, $(A \mid v)_{\bullet,\sim(m+1)} = A$. This proves Proposition 6.134 **(d)**.

(e) Here is an informal proof: Let $p \in \{1, 2, \ldots, n\}$. The matrix $(A \mid v)$ is obtained from the matrix A by attaching the column v at its "right edge". The matrix $(A \mid v)_{\sim p,\bullet}$ is obtained from the matrix $(A \mid v)$ by removing its p-th row. Hence, the matrix $(A \mid v)_{\sim p,\bullet}$ is obtained from the matrix A by first attaching the column v at its "right edge" and then removing the p-th row. On the other hand, the matrix $(A_{\sim p,\bullet} \mid v_{\sim p,\bullet})$ is obtained from the matrix A by first removing the p-th row (this is how we get $A_{\sim p,\bullet}$), and then attaching the column $v_{\sim p,\bullet}$ (this is v with its p-th row removed) at its "right edge". Obviously, the two procedures result in the same matrix, i.e., we have $(A \mid v)_{\sim p,\bullet} = (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})$. Thus, Proposition 6.134 **(e)** is proven.[572]

[572]Here is an outline of a more formal proof: First, we observe that if A is an $n \times m$-matrix, if $p \in \{1, 2, \ldots, n\}$ and if $q \in \{1, 2, \ldots, m\}$, then

$$(A_{\sim p,\bullet})_{\bullet,q} = (A_{\bullet,q})_{\sim p,\bullet}. \qquad (1284)$$

(Indeed, this is a simple fact that is similar to the statements of Proposition 6.130. What it says is that if you remove the p-th row of the matrix A and then take the q-th column of the resulting matrix, then you get the same result as when you first take the q-th column of A and then remove the p-th row of this column.)

Now, let us prove Proposition 6.134 **(e)** formally: Both $(A \mid v)_{\sim p,\bullet}$ and $(A_{\sim p,\bullet} \mid v_{\sim p,\bullet})$ are $(n-1) \times (m+1)$-matrices (since they have one fewer row and one more column than A). We shall now show that

$$\left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} = (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,q} \qquad (1285)$$

for every $q \in \{1, 2, \ldots, m+1\}$.

Proof of (1285): Let $q \in \{1, 2, \ldots, m+1\}$. We must prove (1285).

We can apply (1284) to $m+1$ and $(A \mid v)$ instead of m and A. As a result, we obtain

$$\left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} = \left((A \mid v)_{\bullet,q}\right)_{\sim p,\bullet}. \qquad (1286)$$

We are in one of the following two cases:

Case 1: We have $q \neq m+1$.

Case 2: We have $q = m+1$.

Let us consider Case 1 first. In this case, we have $q \neq m+1$. Hence, $q \in \{1, 2, \ldots, m\}$ (since $q \in \{1, 2, \ldots, m+1\}$). Thus, Proposition 6.134 **(a)** yields $(A \mid v)_{\bullet,q} = A_{\bullet,q}$. Now, (1286) becomes

$$\left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} = \left(\underbrace{(A \mid v)_{\bullet,q}}_{=A_{\bullet,q}}\right)_{\sim p,\bullet} = (A_{\bullet,q})_{\sim p,\bullet}.$$

(f) Let $p \in \{1, 2, \ldots, n\}$. Then, $(A \mid v)$ is an $n \times (m+1)$-matrix. Proposition 6.130 **(c)** (applied to $m+1$, $(A \mid v)$, p and $m+1$ instead of m, A, u and v) yields

$$\left((A \mid v)_{\bullet,\sim(m+1)}\right)_{\sim p,\bullet} = \left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,\sim(m+1)} = (A \mid v)_{\sim p,\sim(m+1)}.$$

Hence,

$$(A \mid v)_{\sim p,\sim(m+1)} = \left(\underbrace{(A \mid v)_{\bullet,\sim(m+1)}}_{\substack{=A \\ \text{(by Proposition 6.134 (d))}}}\right)_{\sim p,\bullet} = A_{\sim p,\bullet}.$$

This proves Proposition 6.134 **(f)**. □

Comparing this with

$$\begin{aligned}(A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,q} &= (A_{\sim p,\bullet})_{\bullet,q} && \left(\begin{array}{c}\text{by Proposition 6.134 (a), applied to} \\ n-1,\ A_{\sim p,\bullet} \text{ and } v_{\sim p,\bullet} \text{ instead of } n,\ A \text{ and } v\end{array}\right) \\ &= (A_{\bullet,q})_{\sim p,\bullet} && \text{(by (1284))},\end{aligned}$$

we obtain $\left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} = (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,q}$. Thus, (1285) is proven in Case 1.

Let us now consider Case 2. In this case, we have $q = m+1$. Hence, $(A \mid v)_{\bullet,q} = (A \mid v)_{\bullet,m+1} = v$ (by Proposition 6.134 **(b)**). Now, (1286) becomes

$$\left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} = \left(\underbrace{(A \mid v)_{\bullet,q}}_{=v}\right)_{\sim p,\bullet} = v_{\sim p,\bullet}.$$

Comparing this with

$$\begin{aligned}(A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,q} &= (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,m+1} && (\text{since } q = m+1) \\ &= v_{\sim p,\bullet} && \left(\begin{array}{c}\text{by Proposition 6.134 (b), applied to} \\ n-1,\ A_{\sim p,\bullet} \text{ and } v_{\sim p,\bullet} \text{ instead of } n,\ A \text{ and } v\end{array}\right),\end{aligned}$$

we obtain $\left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} = (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,q}$. Thus, (1285) is proven in Case 2.

Now we have proven (1285) in each of the two Cases 1 and 2. Thus, the proof of (1285) is complete.

Now, for every $q \in \{1, 2, \ldots, m+1\}$, we have

$$\begin{aligned}&\left(\text{the } q\text{-th column of the matrix } (A \mid v)_{\sim p,\bullet}\right) \\ &= \left((A \mid v)_{\sim p,\bullet}\right)_{\bullet,q} \\ &= (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})_{\bullet,q} && \text{(by (1285))} \\ &= \left(\text{the } q\text{-th column of the matrix } (A_{\sim p,\bullet} \mid v_{\sim p,\bullet})\right).\end{aligned}$$

This shows that the matrices $(A \mid v)_{\sim p,\bullet}$ and $(A_{\sim p,\bullet} \mid v_{\sim p,\bullet})$ are identical (since they are both $(n-1) \times (m+1)$-matrices). Proposition 6.134 **(e)** is thus proven.

Proof of Proposition 6.135. **(a)** Let $v = (v_1, v_2, \ldots, v_n)^T \in \mathbb{K}^{n \times 1}$ be a vector.

Clearly, $(A \mid v)$ is an $n \times n$-matrix (since it is obtained by attaching the column v to the $n \times (n-1)$-matrix A). Write it in the form $(A \mid v) = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Then, every $p \in \{1, 2, \ldots, n\}$ satisfies

$$b_{p,n} = v_p \tag{1287}$$

[573] and

$$(A \mid v)_{\sim p, \sim n} = A_{\sim p, \bullet} \tag{1288}$$

[574].

We have $n \in \{1, 2, \ldots, n\}$ (since n is a positive integer) and $(A \mid v) = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, Theorem 6.82 **(b)** (applied to $(A \mid v)$, $b_{i,j}$ and n instead of A, $a_{i,j}$ and q) yields

$$\begin{aligned}
\det (A \mid v) &= \sum_{p=1}^{n} \underbrace{(-1)^{p+n}}_{= (-1)^{n+p}} \underbrace{b_{p,n}}_{\substack{= v_p \\ \text{(by (1287))}}} \det \left(\underbrace{(A \mid v)_{\sim p, \sim n}}_{\substack{= A_{\sim p, \bullet} \\ \text{(by (1288))}}} \right) \\
&= \sum_{p=1}^{n} (-1)^{n+p} v_p \det \left(A_{\sim p, \bullet} \right) = \sum_{i=1}^{n} (-1)^{n+i} v_i \det \left(A_{\sim i, \bullet} \right)
\end{aligned}$$

---

[573] *Proof of (1287):* We have $(A \mid v) = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus,

$$(\text{the } n\text{-th column of the matrix } (A \mid v)) = \begin{pmatrix} b_{1,n} \\ b_{2,n} \\ \vdots \\ b_{n,n} \end{pmatrix}.$$

Hence,

$$\begin{aligned}
\begin{pmatrix} b_{1,n} \\ b_{2,n} \\ \vdots \\ b_{n,n} \end{pmatrix} &= (\text{the } n\text{-th column of the matrix } (A \mid v)) = (A \mid v)_{\bullet, n} \\
&= (A \mid v)_{\bullet, (n-1)+1} \qquad (\text{since } n = (n-1)+1) \\
&= v \qquad (\text{by Proposition 6.134 } \textbf{(b)}, \text{ applied to } m = n-1) \\
&= (v_1, v_2, \ldots, v_n)^T = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.
\end{aligned}$$

In other words, every $p \in \{1, 2, \ldots, n\}$ satisfies $b_{p,n} = v_p$. This proves (1287).

[574] *Proof of (1288):* Let $p \in \{1, 2, \ldots, n\}$. Proposition 6.134 **(f)** (applied to $m = n-1$) yields $(A \mid v)_{\sim p, \sim ((n-1)+1)} = A_{\sim p, \bullet}$. This rewrites as $(A \mid v)_{\sim p, \sim n} = A_{\sim p, \bullet}$ (since $(n-1)+1 = n$). This proves (1288).

(here, we have renamed the summation index p as i). This proves Proposition 6.135 **(a)**.

(b) Let $p \in \{1, 2, \ldots, n\}$. For any two objects i and j, we define an element $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$. Then, $I_n = \left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of I_n). Now, $\left(I_n\right)_{\bullet,p}$ is the p-th column of the matrix I_n (by the definition of $\left(I_n\right)_{\bullet,p}$). Thus,

$$\begin{aligned}\left(I_n\right)_{\bullet,p} &= \left(\text{the } p\text{-th column of the matrix } I_n\right) \\ &= \begin{pmatrix} \delta_{1,p} \\ \delta_{2,p} \\ \vdots \\ \delta_{n,p} \end{pmatrix} \qquad \left(\text{since } I_n = \left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ &= \left(\delta_{1,p}, \delta_{2,p}, \ldots, \delta_{n,p}\right)^T.\end{aligned}$$

Hence, Proposition 6.135 **(a)** (applied to $\left(I_n\right)_{\bullet,p}$ and $\delta_{i,p}$ instead of v and v_i) yields

$$\begin{aligned}&\det\left(A \mid \left(I_n\right)_{\bullet,p}\right) \\ &= \underbrace{\sum_{i=1}^{n}}_{=\sum_{i\in\{1,2,\ldots,n\}}} (-1)^{n+i}\,\delta_{i,p}\det\left(A_{\sim i,\bullet}\right) = \sum_{i\in\{1,2,\ldots,n\}} (-1)^{n+i}\,\delta_{i,p}\det\left(A_{\sim i,\bullet}\right) \\ &= (-1)^{n+p}\underbrace{\delta_{p,p}}_{\substack{=1 \\ (\text{since } p=p)}}\det\left(A_{\sim p,\bullet}\right) + \sum_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} (-1)^{n+i}\underbrace{\delta_{i,p}}_{\substack{=0 \\ (\text{since } i\neq p)}}\det\left(A_{\sim i,\bullet}\right) \\ &\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } i = p \text{ from the sum,} \\ \text{since } p \in \{1, 2, \ldots, n\}\end{array}\right) \\ &= (-1)^{n+p}\det\left(A_{\sim p,\bullet}\right) + \underbrace{\sum_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} (-1)^{n+i}\,0\det\left(A_{\sim i,\bullet}\right)}_{=0} = (-1)^{n+p}\det\left(A_{\sim p,\bullet}\right).\end{aligned}$$

This proves Proposition 6.135 **(b)**. □

Before we prove Proposition 6.136, let us make a simple observation:

Lemma 7.209. Let $n \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \in \mathbb{K}^{n\times n}$ be an $n \times n$-matrix. Let r and q be two elements of $\{1, 2, \ldots, n\}$. Then,

$$\det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) = (-1)^{n+q}\sum_{p=1}^{n} (-1)^{p+q}\,a_{p,r}\det\left(A_{\sim p,\sim q}\right).$$

Proof of Lemma 7.209. We have $r \in \{1,2,\ldots,n\}$. Hence, $A_{\bullet,r}$ is an $n \times 1$-matrix (since A is an $n \times n$-matrix). Also, $r \in \{1,2,\ldots,n\}$ shows that $1 \leq r \leq n$; hence, $1 \leq n$, so that n is a positive integer.

Also, $q \in \{1,2,\ldots,n\}$. Hence, $A_{\bullet,\sim q}$ is an $n \times (n-1)$-matrix (since A is an $n \times n$-matrix).

Furthermore, the definition of $A_{\bullet,r}$ yields

$$\begin{aligned} A_{\bullet,r} &= (\text{the } r\text{-th column of the matrix } A) \\ &= \begin{pmatrix} a_{1,r} \\ a_{2,r} \\ \vdots \\ a_{n,r} \end{pmatrix} \qquad \left(\text{since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\ &= \left(a_{1,r}, a_{2,r}, \ldots, a_{n,r}\right)^T. \end{aligned}$$

Thus, Proposition 6.135 **(a)** (applied to $A_{\bullet,\sim q}$, $A_{\bullet,r}$ and $a_{i,r}$ instead of A, v and v_i) yields

$$\det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) = \sum_{i=1}^{n} (-1)^{n+i} a_{i,r} \det\left(\left(A_{\bullet,\sim q}\right)_{\sim i,\bullet}\right). \qquad (1289)$$

But every $i \in \{1,2,\ldots,n\}$ satisfies

$$\left(A_{\bullet,\sim q}\right)_{\sim i,\bullet} = A_{\sim i,\sim q} \qquad (1290)$$

[575] and

$$(-1)^{n+i} = (-1)^{n+q} (-1)^{i+q} \qquad (1291)$$

[576]. Now, (1289) becomes

$$\begin{aligned} \det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) &= \sum_{i=1}^{n} \underbrace{(-1)^{n+i}}_{\substack{=(-1)^{n+q}(-1)^{i+q} \\ \text{(by (1291))}}} a_{i,r} \det\left(\underbrace{\left(A_{\bullet,\sim q}\right)_{\sim i,\bullet}}_{\substack{=A_{\sim i,\sim q} \\ \text{(by (1290))}}}\right) \\ &= \sum_{i=1}^{n} (-1)^{n+q} (-1)^{i+q} a_{i,r} \det\left(A_{\sim i,\sim q}\right) \\ &= (-1)^{n+q} \sum_{i=1}^{n} (-1)^{i+q} a_{i,r} \det\left(A_{\sim i,\sim q}\right) \\ &= (-1)^{n+q} \sum_{p=1}^{n} (-1)^{p+q} a_{p,r} \det\left(A_{\sim p,\sim q}\right) \\ &\qquad (\text{here, we have renamed the summation index } i \text{ as } p). \end{aligned}$$

[575] *Proof of (1290):* Let $i \in \{1,2,\ldots,n\}$. Then, Proposition 6.130 **(c)** (applied to n, i and q instead of m, u and v) yields $\left(A_{\bullet,\sim q}\right)_{\sim i,\bullet} = \left(A_{\sim i,\bullet}\right)_{\bullet,\sim q} = A_{\sim i,\sim q}$. This proves (1290).

[576] *Proof of (1291):* Let $i \in \{1,2,\ldots,n\}$. Then, $(-1)^{n+q} (-1)^{i+q} = (-1)^{(n+q)+(i+q)} = (-1)^{n+i}$ (since $(n+q)+(i+q) = n+i+2q \equiv n+i \mod 2$). This proves (1291).

This proves Lemma 7.209. □

Proof of Proposition 6.136. **(a)** Proposition 6.136 **(a)** simply says that if we remove the n-th column from the matrix A, and then reattach this column back to the matrix (at its right edge), then we get the original matrix A back. This should be completely obvious[577].

(b) Write the matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. (This is possible since A is an $n\times n$-matrix.) Lemma 7.209 (applied to $r = q$) yields

$$\det\left(A_{\bullet,\sim q} \mid A_{\bullet,q}\right) = (-1)^{n+q} \underbrace{\sum_{p=1}^{n} (-1)^{p+q} a_{p,q} \det\left(A_{\sim p,\sim q}\right)}_{\substack{=\det A \\ \text{(by Theorem 6.82 (b))}}} = (-1)^{n+q} \det A.$$

This proves Proposition 6.136 **(b)**.

[*Remark:* Let me outline an alternative proof of Proposition 6.136 **(b)**: Let $q \in \{1,2,\ldots,n\}$. The matrix $\left(A_{\bullet,\sim q} \mid A_{\bullet,q}\right)$ is obtained from the matrix A by removing the q-th column and then reattaching this column to the right end of the matrix. This procedure can be replaced by the following procedure, which clearly leads to the same result:

[577]Nevertheless, let me give a formal proof of Proposition 6.136 **(a)** as well:

Assume that $n > 0$. Thus, $n \in \{1,2,\ldots,n\}$. Hence, $A_{\bullet,\sim n}$ is a well-defined $n\times(n-1)$-matrix, and $A_{\bullet,n}$ is a well-defined $n\times 1$-matrix. Therefore, $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)$ is an $n\times n$-matrix.

We shall now show that

$$A_{\bullet,q} = \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q} \qquad (1292)$$

for each $q \in \{1,2,\ldots,n\}$.

Proof of (1292): Let $q \in \{1,2,\ldots,n\}$. We must prove the equality (1292). We are in one of the following two cases:

Case 1: We have $q \neq n$.

Case 2: We have $q = n$.

Let us first consider Case 1. In this case, we have $q \neq n$. Combining $q \in \{1,2,\ldots,n\}$ with $q \neq n$, we obtain $q \in \{1,2,\ldots,n\} \setminus \{n\} = \{1,2,\ldots,n-1\}$. Thus, Proposition 6.134 **(a)** (applied to $n-1$, $A_{\bullet,\sim n}$ and $A_{\bullet,n}$ instead of m, A and v) yields $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q} = \left(A_{\bullet,\sim n}\right)_{\bullet,q}$.

But Proposition 6.130 **(d)** (applied to n, n and q instead of m, v and w) yields $\left(A_{\bullet,\sim n}\right)_{\bullet,q} = A_{\bullet,q}$. Hence, $A_{\bullet,q} = \left(A_{\bullet,\sim n}\right)_{\bullet,q} = \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q}$ (since $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q} = \left(A_{\bullet,\sim n}\right)_{\bullet,q}$). Hence, (1292) is proven in Case 1.

Let us now consider Case 2. In this case, we have $q = n$. But Proposition 6.134 **(b)** (applied to $n-1$, $A_{\bullet,\sim n}$ and $A_{\bullet,n}$ instead of m, A and v) yields $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,(n-1)+1} = A_{\bullet,n}$. Thus, $A_{\bullet,n} = \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,(n-1)+1} = \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q}$ (since $(n-1)+1 = n = q$). Now, $q = n$, so that $A_{\bullet,q} = A_{\bullet,n} = \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q}$. Hence, (1292) is proven in Case 2.

Now, (1292) is proven in each of the two Cases 1 and 2. This completes the proof of (1292).

Now we know that A and $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)$ are two $n\times n$-matrices, and that every $q \in \{1,2,\ldots,n\}$ satisfies

$$\begin{aligned}(\text{the } q\text{-th column of the matrix } A) &= A_{\bullet,q} = \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)_{\bullet,q} \qquad \text{(by (1292))} \\ &= \left(\text{the } q\text{-th column of the matrix } \left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)\right).\end{aligned}$$

Hence, the matrices A and $\left(A_{\bullet,\sim n} \mid A_{\bullet,n}\right)$ are identical. This proves Proposition 6.136 **(a)**.

- Switch the q-th column of A with the $(q+1)$-th column;
- then switch the $(q+1)$-th column of the resulting matrix with the $(q+2)$-th column;
- then switch the $(q+2)$-th column of the resulting matrix with the $(q+3)$-th column;
- and so on, finally switching the $(n-1)$-st column of the matrix with the n-th column.

But this latter procedure is a sequence of $n-q$ switches of two columns. Each such switch multiplies the determinant of the matrix by -1 (according to Exercise 6.7 **(b)**). Thus, the whole procedure multiplies the determinant of the matrix by $(-1)^{n-q} = (-1)^{n+q}$ (since $n-q \equiv n+q \mod 2$). Since this procedure takes the matrix A to the matrix $\left(A_{\bullet,\sim q} \mid A_{\bullet,q}\right)$, we thus conclude that $\det\left(A_{\bullet,\sim q} \mid A_{\bullet,q}\right) = (-1)^{n+q} \det A$. This proves Proposition 6.136 **(b)** again.]

(c) Let r and q be two elements of $\{1,2,\ldots,n\}$ satisfying $r \neq q$. We must show that $\det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) = 0$.

Write the matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. (This is possible since A is an $n\times n$-matrix.) Lemma 7.209 yields

$$\det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) = (-1)^{n+q} \underbrace{\sum_{p=1}^{n} (-1)^{p+q} a_{p,r} \det\left(A_{\sim p,\sim q}\right)}_{\substack{=0 \\ \text{(by Proposition 6.96 (b)} \\ \text{(since } q\neq r\text{))}}} = 0.$$

This proves Proposition 6.136 **(c)**.

[*Remark:* Let me outline an alternative proof of Proposition 6.136 **(c)**: Let r and q be two elements of $\{1,2,\ldots,n\}$ satisfying $r \neq q$. Then, $A_{\bullet,r}$ is one of the columns of the matrix $A_{\bullet,\sim q}$ (since the r-th column of the matrix A is not lost when the q-th column is removed). Hence, the column vector $A_{\bullet,r}$ appears twice as a column in the matrix $\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right)$ (namely, once as one of the columns of $A_{\bullet,\sim q}$, and another time as the attached column). Therefore, the matrix $\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right)$ has two equal columns. Exercise 6.7 **(f)** thus shows that $\det\left(A_{\bullet,\sim q} \mid A_{\bullet,r}\right) = 0$. This proves Proposition 6.136 **(c)** again.]

(d) Let $p \in \{1,2,\ldots,n\}$ and $q \in \{1,2,\ldots,n\}$. Then, Proposition 6.130 **(c)** (applied to $u = p$ and $v = q$) yields

$$\left(A_{\bullet,\sim q}\right)_{\sim p,\bullet} = \left(A_{\sim p,\bullet}\right)_{\bullet,\sim q} = A_{\sim p,\sim q}.$$

But A is an $n\times n$-matrix. Hence, $A_{\bullet,\sim q}$ is an $n\times(n-1)$-matrix. Moreover, $p \in \{1,2,\ldots,n\}$, so that $1 \leq p \leq n$ and thus $n \geq 1$; hence, n is a positive integer.

Proposition 6.135 **(b)** (applied to $A_{\bullet,\sim q}$ instead of A) thus yields

$$\det\left(A_{\bullet,\sim q} \mid (I_n)_{\bullet,p}\right) = (-1)^{n+p} \det\left(\underbrace{\left(A_{\bullet,\sim q}\right)_{\sim p,\bullet}}_{=A_{\sim p,\sim q}}\right) = (-1)^{n+p} \det\left(A_{\sim p,\sim q}\right).$$

This proves Proposition 6.136 **(d)**.

(e) Let u and v be two elements of $\{1,2,\ldots,n\}$ satisfying $u < v$. Let r be an element of $\{1,2,\ldots,n-1\}$ satisfying $r \neq u$. We must show that $\det\left(A_{\bullet,\sim u} \mid (A_{\bullet,\sim v})_{\bullet,r}\right) = 0$.

We are in one of the following two cases:

Case 1: We have $r < v$.

Case 2: We have $r \geq v$.

Let us first consider Case 1. In this case, we have $r < v$. Thus, $r \in \{1,2,\ldots,v-1\}$. Hence, Proposition 6.130 **(d)** (applied to $m = n$ and $w = r$) yields $(A_{\bullet,\sim v})_{\bullet,r} = A_{\bullet,r}$. Hence,

$$\det\left(A_{\bullet,\sim u} \mid \underbrace{(A_{\bullet,\sim v})_{\bullet,r}}_{=A_{\bullet,r}}\right) = \det\left(A_{\bullet,\sim u} \mid A_{\bullet,r}\right) = 0$$

(by Proposition 6.136 **(c)**, applied to $q = u$). Hence, Proposition 6.136 **(e)** is proven in Case 1.

Let us now consider Case 2. In this case, we have $r \geq v$. Hence, $r \in \{v, v+1, \ldots, n-1\}$ (since $r \in \{1,2,\ldots,n-1\}$), so that $r+1 \in \{v+1, v+2, \ldots, n\} \subseteq \{1,2,\ldots,n\}$. Furthermore, $r+1 > r \geq v > u$ (since $u < v$) and thus $r+1 \neq u$. Hence, Proposition 6.136 **(c)** (applied to $r+1$ and u instead of r and q) yields $\det\left(A_{\bullet,\sim u} \mid A_{\bullet,r+1}\right) = 0$.

But recall that $r \in \{v, v+1, \ldots, n-1\}$. Thus, Proposition 6.130 **(e)** (applied to $m = n$ and $w = r$) yields $(A_{\bullet,\sim v})_{\bullet,r} = A_{\bullet,r+1}$. Hence,

$$\det\left(A_{\bullet,\sim u} \mid \underbrace{(A_{\bullet,\sim v})_{\bullet,r}}_{=A_{\bullet,r+1}}\right) = \det\left(A_{\bullet,\sim u} \mid A_{\bullet,r+1}\right) = 0.$$

Hence, Proposition 6.136 **(e)** is proven in Case 2.

We have now proven Proposition 6.136 **(e)** in each of the two Cases 1 and 2. Hence, Proposition 6.136 **(e)** always holds.

(f) Let u and v be two elements of $\{1,2,\ldots,n\}$ satisfying $u < v$. Then, $u < v$, so that $u \in \{1,2,\ldots,v-1\}$. Hence, Proposition 6.130 **(d)** (applied to $m = n$ and $w = u$) yields $(A_{\bullet,\sim v})_{\bullet,u} = A_{\bullet,u}$. Thus,

$$\det\left(A_{\bullet,\sim u} \mid \underbrace{(A_{\bullet,\sim v})_{\bullet,u}}_{=A_{\bullet,u}}\right) = \det\left(A_{\bullet,\sim u} \mid A_{\bullet,u}\right) = (-1)^{n+u} \det A$$

(by Proposition 6.136 **(b)**, applied to $q = u$). Thus,

$$(-1)^u \underbrace{\det\left(A_{\bullet,\sim u} \mid (A_{\bullet,\sim v})_{\bullet,u}\right)}_{=(-1)^{n+u}\det A} = \underbrace{(-1)^u (-1)^{n+u}}_{\substack{=(-1)^{u+(n+u)}=(-1)^n \\ (\text{since } u+(n+u)=2u+n\equiv n \bmod 2)}} \det A = (-1)^n \det A.$$

This proves Proposition 6.136 **(f)**. □

Solution to Exercise 6.40. We have proven Proposition 6.134, Proposition 6.135 and Proposition 6.136. Thus, Exercise 6.40 is solved. □

7.109. Solution to Exercise 6.41

Solution to Exercise 6.41. We have $u < v$ and $u < w$. Define an $n \times n$-matrix $C \in \mathbb{K}^{n\times n}$ as in Lemma 6.139.

Lemma 6.139 **(a)** yields

$$\begin{aligned} \det\left(C_{\sim v,\sim q}\right) &= -(-1)^{n+u} \underbrace{\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)}_{\substack{=\beta_{u,v} \\ (\text{since } \beta_{u,v}=\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{v},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) \\ (\text{by the definition of } \beta_{u,v}))}} \\ &= -(-1)^{n+u}\beta_{u,v}. \end{aligned} \tag{1293}$$

Lemma 6.139 **(a)** (applied to w instead of v) yields

$$\begin{aligned} \det\left(C_{\sim w,\sim q}\right) &= -(-1)^{n+u} \underbrace{\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{w},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)}_{\substack{=\beta_{u,w} \\ (\text{since } \beta_{u,w}=\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{u},\ldots,\widehat{w},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) \\ (\text{by the definition of } \beta_{u,w}))}} \\ &= -(-1)^{n+u}\beta_{u,w}. \end{aligned} \tag{1294}$$

Moreover,

$$\operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C = \operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n}\left(B_{\bullet,\sim q}\right) \tag{1295}$$

[578]. Taking determinants on both sides of this equality, we obtain

$$\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C\right) = \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n}\left(B_{\bullet,\sim q}\right)\right) = \beta_{v,w} \tag{1297}$$

[578] *Proof of (1295):* From $C = \left(B \mid (I_n)_{\bullet,u}\right)$, we obtain

$$\begin{aligned} C_{\bullet,\sim n} &= \left(B \mid (I_n)_{\bullet,u}\right)_{\bullet,\sim n} = \left(B \mid (I_n)_{\bullet,u}\right)_{\bullet,\sim((n-1)+1)} \qquad (\text{since } n = (n-1)+1) \\ &= B \end{aligned}$$

(by Proposition 6.134 **(d)**, applied to $n-1$, B and $(I_n)_{\bullet,u}$ instead of m, A and v).

Proposition 6.79 **(d)** (applied to n, C, $n-2$, $(1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n)$, $n-2$ and (since $\beta_{v,w} = \det\left(\operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n}\left(B_{\bullet,\sim q}\right)\right)$ (by the definition of $\beta_{v,w}$)).

Lemma 6.139 **(c)** yields $C_{\sim v,\sim n} = B_{\sim v,\bullet}$. Thus,

$$\det\left(\underbrace{C_{\sim v,\sim n}}_{=B_{\sim v,\bullet}}\right) = \det\left(B_{\sim v,\bullet}\right) = \alpha_v \tag{1298}$$

(since $\alpha_v = \det\left(B_{\sim v,\bullet}\right)$ (by the definition of α_v)).

Lemma 6.139 **(c)** (applied to w instead of v) yields $C_{\sim w,\sim n} = B_{\sim w,\bullet}$. Thus,

$$\det\left(\underbrace{C_{\sim w,\sim n}}_{=B_{\sim w,\bullet}}\right) = \det\left(B_{\sim w,\bullet}\right) = \alpha_w \tag{1299}$$

(since $\alpha_w = \det\left(B_{\sim w,\bullet}\right)$ (by the definition of α_w)).

Lemma 6.139 **(f)** yields

$$\det C = (-1)^{n+u} \underbrace{\det\left(B_{\sim u,\bullet}\right)}_{\substack{=\alpha_u \\ (\text{since } \alpha_u = \det(B_{\sim u,\bullet}) \\ (\text{by the definition of } \alpha_u))}} = (-1)^{n+u}\alpha_u. \tag{1300}$$

Now, $q \in \{1,2,\ldots,n-1\}$, so that $q \leq n-1 < n$. Also, $n \in \{1,2,\ldots,n\}$ (since n is a positive integer) and $q \in \{1,2,\ldots,n-1\} \subseteq \{1,2,\ldots,n\}$. Hence, Theorem

---

$(1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n)$ instead of m, A, u, $(i_1,i_2,\ldots,i_u)$, v and $(j_1,j_2,\ldots,j_v)$) yields

$$\begin{aligned}\operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C &= \operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n}\left(\operatorname{cols}_{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n} C\right) \\ &= \operatorname{cols}_{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}\left(\operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C\right).\end{aligned} \tag{1296}$$

But n is a positive integer; hence, $n \in \{1,2,\ldots,n\}$. Also, $q \in \{1,2,\ldots,n-1\}$. Thus, Proposition 6.130 **(h)** (applied to n, C, n and q instead of m, A, v and w) yields

$$\left(C_{\bullet,\sim n}\right)_{\bullet,\sim q} = \operatorname{cols}_{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n} C.$$

Hence,

$$\operatorname{cols}_{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n} C = \left(\underbrace{C_{\bullet,\sim n}}_{=B}\right)_{\bullet,\sim q} = B_{\bullet,\sim q}.$$

Hence, (1296) becomes

$$\begin{aligned}\operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C &= \operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n}\left(\underbrace{\operatorname{cols}_{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n} C}_{=B_{\bullet,\sim q}}\right) \\ &= \operatorname{rows}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n}\left(B_{\bullet,\sim q}\right).\end{aligned}$$

This proves (1295).

6.126 (applied to C, v, w, q and n instead of A, p, q, u and v) yields

$$\begin{aligned}
&\det C \cdot \det\left(\operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C\right) \\
&= \underbrace{\det\left(C_{\sim v,\sim q}\right)}_{\substack{=-(-1)^{n+u}\beta_{u,v} \\ \text{(by (1293))}}} \cdot \underbrace{\det\left(C_{\sim w,\sim n}\right)}_{\substack{=\alpha_w \\ \text{(by (1299))}}} - \underbrace{\det\left(C_{\sim v,\sim n}\right)}_{\substack{=\alpha_v \\ \text{(by (1298))}}} \cdot \underbrace{\det\left(C_{\sim w,\sim q}\right)}_{\substack{=-(-1)^{n+u}\beta_{u,w} \\ \text{(by (1294))}}} \\
&= \left(-(-1)^{n+u}\beta_{u,v}\right) \cdot \alpha_w - \alpha_v \cdot \left(-(-1)^{n+u}\beta_{u,w}\right) \\
&= -(-1)^{n+u}\beta_{u,v} \cdot \alpha_w + \alpha_v \cdot (-1)^{n+u}\beta_{u,w} \\
&= \alpha_v \cdot (-1)^{n+u}\beta_{u,w} - (-1)^{n+u}\beta_{u,v} \cdot \alpha_w = (-1)^{n+u}\left(\alpha_v\beta_{u,w} - \beta_{u,v}\alpha_w\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
(-1)^{n+u}\left(\alpha_v\beta_{u,w} - \beta_{u,v}\alpha_w\right) &= \underbrace{\det C}_{\substack{=(-1)^{n+u}\alpha_u \\ \text{(by (1300))}}} \cdot \underbrace{\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{q},\ldots,\widehat{n},\ldots,n}_{1,2,\ldots,\widehat{v},\ldots,\widehat{w},\ldots,n} C\right)}_{\substack{=\beta_{v,w} \\ \text{(by (1297))}}} \\
&= (-1)^{n+u}\alpha_u\beta_{v,w}.
\end{aligned}$$

Multiplying both sides of this equality by $(-1)^{n+u}$, we find

$$(-1)^{n+u}(-1)^{n+u}\left(\alpha_v\beta_{u,w} - \beta_{u,v}\alpha_w\right) = \underbrace{(-1)^{n+u}(-1)^{n+u}}_{\substack{=(-1)^{(n+u)+(n+u)}=1 \\ \text{(since } (n+u)+(n+u)=2(n+u) \text{ is even)}}} \alpha_u\beta_{v,w} = \alpha_u\beta_{v,w}.$$

Hence,

$$\begin{aligned}
\alpha_u\beta_{v,w} &= \underbrace{(-1)^{n+u}(-1)^{n+u}}_{\substack{=(-1)^{(n+u)+(n+u)}=1 \\ \text{(since } (n+u)+(n+u)=2(n+u) \text{ is even)}}} \left(\alpha_v\beta_{u,w} - \beta_{u,v}\alpha_w\right) = \alpha_v\beta_{u,w} - \beta_{u,v}\alpha_w \\
&= \alpha_v\beta_{u,w} - \alpha_w\beta_{u,v}.
\end{aligned}$$

Adding $\alpha_w\beta_{u,v}$ to both sides of this equality, we obtain $\alpha_u\beta_{v,w} + \alpha_w\beta_{u,v} = \alpha_v\beta_{u,w}$. This solves Exercise 6.41. □

7.110. Solution to Exercise 6.42

Before we solve Exercise 6.42], let us state a variant of Lemma 7.168 for the case when A is alternating:

Lemma 7.210. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an alternating $n \times n$-matrix. Let $S = \left(s_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ be an $n \times m$-matrix. Then,

$$S^T A S = \left(\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2; \\ k<\ell}} \left(s_{k,i}s_{\ell,j} - s_{\ell,i}s_{k,j}\right) a_{k,\ell}\right)_{1\leq i\leq m,\ 1\leq j\leq m}.$$

Proof of Lemma 7.210. Every $(i,j) \in \{1,2,\ldots,m\}^2$ satisfies

$$\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2} s_{k,i}s_{\ell,j}a_{k,\ell} = \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} \left(s_{k,i}s_{\ell,j} - s_{\ell,i}s_{k,j}\right) a_{k,\ell}. \tag{1301}$$

[*Proof of (1301):* Let $(i,j) \in \{1,2,\ldots,m\}^2$. Then,

$$\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k=\ell}} s_{k,i}s_{\ell,j} \underbrace{a_{k,\ell}}_{\substack{=a_{\ell,\ell}\\ \text{(since } k=\ell)}} = \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k=\ell}} s_{k,i}s_{\ell,j} \underbrace{a_{\ell,\ell}}_{\substack{=0\\ \text{(by Lemma 7.169 \textbf{(a)}}\\ \text{(applied to } i=\ell))}}$$
$$= \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k=\ell}} s_{k,i}s_{\ell,j}0 = 0. \tag{1302}$$

Also,

$$\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k>\ell}} s_{k,i}s_{\ell,j} \underbrace{a_{k,\ell}}_{\substack{=-a_{\ell,k}\\ \text{(by Lemma 7.169 \textbf{(b)}}\\ \text{(applied to } (i,j)=(k,\ell)))}}$$
$$= -\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k>\ell}} s_{k,i}s_{\ell,j}a_{\ell,k} = -\underbrace{\sum_{\substack{(\ell,k)\in\{1,2,\ldots,n\}^2;\\ \ell>k}}}_{=\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ \ell>k}} = \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}}} s_{\ell,i}s_{k,j}a_{k,\ell}$$
$$\left(\text{here, we have renamed the summation index } (k,\ell) \text{ as } (\ell,k)\right)$$
$$= -\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} s_{\ell,i}s_{k,j}a_{k,\ell}. \tag{1303}$$

Now,

$$\begin{aligned}
&\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2} s_{k,i}s_{\ell,j}a_{k,\ell}\\
&= \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} s_{k,i}s_{\ell,j}a_{k,\ell} + \underbrace{\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k=\ell}} s_{k,i}s_{\ell,j}a_{k,\ell}}_{\substack{=0\\ \text{(by (1302))}}}\\
&\qquad + \underbrace{\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k>\ell}} s_{k,i}s_{\ell,j}a_{k,\ell}}_{\substack{=-\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} s_{\ell,i}s_{k,j}a_{k,\ell}\\ \text{(by (1303))}}}\\
&\qquad\left(\begin{array}{c}\text{since each } (k,\ell)\in\{1,2,\ldots,n\}^2 \text{ satisfies exactly one of}\\ \text{the three assertions } (k<\ell),\ (k=\ell) \text{ and } (k>\ell)\end{array}\right)\\
&= \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} s_{k,i}s_{\ell,j}a_{k,\ell} - \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} s_{\ell,i}s_{k,j}a_{k,\ell}\\
&= \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} \left(s_{k,i}s_{\ell,j} - s_{\ell,i}s_{k,j}\right) a_{k,\ell}.
\end{aligned}$$

This proves (1301).]

Lemma 7.168 yields

$$\begin{aligned}
S^TAS &= \left(\underbrace{\sum_{(k,\ell)\in\{1,2,\ldots,n\}^2} s_{k,i}s_{\ell,j}a_{k,\ell}}_{\substack{=\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} \left(s_{k,i}s_{\ell,j}-s_{\ell,i}s_{k,j}\right)a_{k,\ell}\\ \text{(by (1301))}}}\right)_{1\le i\le m,\ 1\le j\le m}\\
&= \left(\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} \left(s_{k,i}s_{\ell,j} - s_{\ell,i}s_{k,j}\right) a_{k,\ell}\right)_{1\le i\le m,\ 1\le j\le m}.
\end{aligned}$$

Thus, Lemma 7.210. □

Another simple lemma that we will use is the following:

Lemma 7.211. Let $n \in \mathbb{N}$. Let $B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an alternating $n \times n$-matrix. Let $c \in \mathbb{K}$. Assume that for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$, the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c. Then, each entry of B is a multiple of c.

Proof of Lemma 7.211. We have assumed that the following holds:

> *Fact 1:* For every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$, the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c.

From this, it is easy to derive the following claim:

> *Fact 2:* For every $(i,j) \in \{1,2,\ldots,n\}^2$, the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c.

[*Proof of Fact 2:* Let $(i,j) \in \{1,2,\ldots,n\}^2$. We must show that the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c.

We are in one of the following three cases:

Case 1: We have $i < j$.

Case 2: We have $i = j$.

Case 3: We have $i > j$.

Let us first consider Case 1. In this case, we have $i < j$. Hence, the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c (by Fact 1). Thus, Fact 2 is proven in Case 1.

Let us next consider Case 2. In this case, we have $i = j$. Hence, $j = i$. Therefore, $b_{i,j} = b_{i,i} = 0$ (by Lemma 7.169 **(a)** (applied to B and $b_{u,v}$ instead of A and $a_{u,v}$)). Thus, the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c (since the element 0 of $\mathbb{K}$ is a multiple of c). Hence, Fact 2 is proven in Case 2.

Let us finally consider Case 3. In this case, we have $i > j$. Thus, $j < i$. Hence, Fact 1 (applied to (j,i) instead of (i,j)) shows that the element $b_{j,i}$ of $\mathbb{K}$ is a multiple of c. In other words, $b_{j,i} = dc$ for some $d \in \mathbb{K}$. Consider this d. But Lemma 7.169 **(b)** (applied to B and $b_{u,v}$ instead of A and $a_{u,v}$) yields $b_{i,j} = -\underbrace{b_{j,i}}_{=dc} = -dc = (-d)\,c$. Hence, the element $b_{i,j}$ of $\mathbb{K}$ is a multiple of c. Thus, Fact 2 is proven in Case 3.

We have now proven Fact 2 in each of the three Cases 1, 2 and 3. Since these three Cases cover all possibilities, we thus conclude that Fact 2 always holds.]

Notice that $B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus, for every $(i,j) \in \{1,2,\ldots,n\}^2$, we have

$$(\text{the } (i,j)\text{-th entry of } B) = b_{i,j}. \tag{1304}$$

We now need to prove that each entry of B is a multiple of c. In other words, we need to show that, for every $(i,j) \in \{1,2,\ldots,n\}^2$, the (i,j)-th entry of B is a multiple of c. So let us fix $(i,j) \in \{1,2,\ldots,n\}^2$. Then, Fact 2 shows that $b_{i,j}$ is a multiple of c. In light of (1304), this rewrites as follows: The (i,j)-th entry of B is a multiple of c. This is precisely what we wanted to prove. Thus, the proof of Lemma 7.211 is complete. □

Solution to Exercise 6.42. Theorem 6.126 leads to the following fact: If i, j, k and ℓ are four elements of $\{1,2,\ldots,n\}$ such that $i<j$ and $k<\ell$, then

$$\begin{aligned}&\det S\cdot\det\left(\operatorname{sub}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}S\right)\\&=\det\left(S_{\sim i,\sim k}\right)\cdot\det\left(S_{\sim j,\sim\ell}\right)-\det\left(S_{\sim i,\sim\ell}\right)\cdot\det\left(S_{\sim j,\sim k}\right)\end{aligned}\qquad(1305)$$

[579].

For each $(i,j)\in\{1,2,\ldots,n\}^2$, define an element $s_{i,j}$ of $\mathbb{K}$ by

$$s_{i,j}=(-1)^{i+j}\det\left(S_{\sim j,\sim i}\right).\qquad(1306)$$

(Notice that these elements $s_{i,j}$ are **not** supposed to be the entries of S, despite the notation!)

The definition of $\operatorname{adj}S$ yields

$$\operatorname{adj}S=\left(\underbrace{(-1)^{i+j}\det\left(S_{\sim j,\sim i}\right)}_{\substack{=s_{i,j}\\ \text{(by (1306))}}}\right)_{1\leq i\leq n,\ 1\leq j\leq n}=\left(s_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Thus, Lemma 7.210 (applied to n and $\operatorname{adj}S$ instead of m and S) yields

$$(\operatorname{adj}S)^T A(\operatorname{adj}S)=\left(\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}}\left(s_{k,i}s_{\ell,j}-s_{\ell,i}s_{k,j}\right)a_{k,\ell}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.\qquad(1307)$$

Moreover, Exercise 6.23 (applied to n and $\operatorname{adj}S$ instead of m and S) yields that the $n\times n$-matrix $(\operatorname{adj}S)^T A(\operatorname{adj}S)$ is alternating.

[579] *Proof of (1305):* Let i, j, k and ℓ be four elements of $\{1,2,\ldots,n\}$ such that $i<j$ and $k<\ell$. We have $i\geq 1$ (since $i\in\{1,2,\ldots,n\}$) and $j\leq n$ (since $j\in\{1,2,\ldots,n\}$). But $i<j$ and thus $i\leq j-1$ (since i and j are integers). Hence, $1\leq i\leq\underbrace{j}_{\leq n}-1\leq n-1$, so that $n-1\geq 1$ and thus $n\geq 2$. Hence, Theorem 6.126 (applied to S, i, j, k and ℓ instead of A, p, q, u and v) yields

$$\det S\cdot\det\left(\operatorname{sub}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}S\right)=\det\left(S_{\sim i,\sim k}\right)\cdot\det\left(S_{\sim j,\sim\ell}\right)-\det\left(S_{\sim i,\sim\ell}\right)\cdot\det\left(S_{\sim j,\sim k}\right).$$

This proves (1305).

But if i, j, k and ℓ are four elements of $\{1,2,\ldots,n\}$ such that $i<j$ and $k<\ell$, then

$$\underbrace{s_{k,i}}_{\substack{=(-1)^{k+i}\det\left(S_{\sim i,\sim k}\right)\\ \text{(by the definition of } s_{k,i})}} \underbrace{s_{\ell,j}}_{\substack{=(-1)^{\ell+j}\det\left(S_{\sim j,\sim \ell}\right)\\ \text{(by the definition of } s_{\ell,j})}} - \underbrace{s_{\ell,i}}_{\substack{=(-1)^{\ell+i}\det\left(S_{\sim i,\sim \ell}\right)\\ \text{(by the definition of } s_{\ell,i})}} \underbrace{s_{k,j}}_{\substack{=(-1)^{k+j}\det\left(S_{\sim j,\sim k}\right)\\ \text{(by the definition of } s_{k,j})}}$$
$$= \underbrace{(-1)^{k+i}\det\left(S_{\sim i,\sim k}\right)\cdot(-1)^{\ell+j}\det\left(S_{\sim j,\sim \ell}\right)}_{=(-1)^{k+i}(-1)^{\ell+j}\det\left(S_{\sim i,\sim k}\right)\cdot\det\left(S_{\sim j,\sim \ell}\right)} - \underbrace{(-1)^{\ell+i}\det\left(S_{\sim i,\sim \ell}\right)\cdot(-1)^{k+j}\det\left(S_{\sim j,\sim k}\right)}_{=(-1)^{\ell+i}(-1)^{k+j}\det\left(S_{\sim i,\sim \ell}\right)\cdot\det\left(S_{\sim j,\sim k}\right)}$$
$$= \underbrace{(-1)^{k+i}(-1)^{\ell+j}}_{\substack{=(-1)^{(k+i)+(\ell+j)}=(-1)^{k+\ell+i+j}\\ \text{(since } (k+i)+(\ell+j)=k+\ell+i+j)}} \det\left(S_{\sim i,\sim k}\right)\cdot\det\left(S_{\sim j,\sim \ell}\right)$$
$$- \underbrace{(-1)^{\ell+i}(-1)^{k+j}}_{\substack{=(-1)^{(\ell+i)+(k+j)}=(-1)^{k+\ell+i+j}\\ \text{(since } (\ell+i)+(k+j)=k+\ell+i+j)}} \det\left(S_{\sim i,\sim \ell}\right)\cdot\det\left(S_{\sim j,\sim k}\right)$$
$$= (-1)^{k+\ell+i+j}\det\left(S_{\sim i,\sim k}\right)\cdot\det\left(S_{\sim j,\sim \ell}\right) - (-1)^{k+\ell+i+j}\det\left(S_{\sim i,\sim \ell}\right)\cdot\det\left(S_{\sim j,\sim k}\right)$$
$$= (-1)^{k+\ell+i+j}\underbrace{\left(\det\left(S_{\sim i,\sim k}\right)\cdot\det\left(S_{\sim j,\sim \ell}\right) - \det\left(S_{\sim i,\sim \ell}\right)\cdot\det\left(S_{\sim j,\sim k}\right)\right)}_{\substack{=\det S\cdot\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n} S\right)\\ \text{(by (1305))}}}$$
$$= (-1)^{k+\ell+i+j}\det S\cdot\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n} S\right). \qquad (1308)$$

Hence, for every $(i,j)\in\{1,2,\ldots,n\}^2$ satisfying $i<j$, the element $\sum\limits_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} \left(s_{k,i}s_{\ell,j}-s_{\ell,i}s_{k,j}\right)a_{k,\ell}$ of $\mathbb{K}$ is a multiple of $\det S$ [580]. Thus, Lemma

[580] *Proof.* Let $(i,j)\in\{1,2,\ldots,n\}^2$ be such that $i<j$. From $(i,j)\in\{1,2,\ldots,n\}^2$, we obtain $i\in\{1,2,\ldots,n\}$ and $j\in\{1,2,\ldots,n\}$. Now,

$$\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} \underbrace{\left(s_{k,i}s_{\ell,j}-s_{\ell,i}s_{k,j}\right)}_{\substack{=(-1)^{k+\ell+i+j}\det S\cdot\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n} S\right)\\ \text{(by (1308))}}} a_{k,\ell}$$
$$= \sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} (-1)^{k+\ell+i+j}\det S\cdot\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n} S\right)a_{k,\ell}$$
$$= \left(\sum_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} (-1)^{k+\ell+i+j}\det\left(\operatorname{sub}^{1,2,\ldots,\widehat{k},\ldots,\widehat{\ell},\ldots,n}_{1,2,\ldots,\widehat{i},\ldots,\widehat{j},\ldots,n} S\right)a_{k,\ell}\right)\det S$$

is clearly a multiple of $\det S$. Qed.

7.211 (applied to $B = (\operatorname{adj} S)^T A (\operatorname{adj} S)$, $b_{i,j} = \sum\limits_{\substack{(k,\ell)\in\{1,2,\ldots,n\}^2;\\ k<\ell}} (s_{k,i}s_{\ell,j} - s_{\ell,i}s_{k,j}) a_{k,\ell}$ and $c = \det S$) shows that each entry of the matrix $(\operatorname{adj} S)^T A (\operatorname{adj} S)$ is a multiple of $\det S$ (because the matrix $(\operatorname{adj} S)^T A (\operatorname{adj} S)$ is alternating and satisfies (1307)). This solves Exercise 6.42. □

7.111. Solution to Exercise 6.43

Second proof of Proposition 6.137. Recall that I_m denotes the $m \times m$ identity matrix for each $m \in \mathbb{N}$. Thus, I_n is the $n \times n$ identity matrix. Hence, $(I_n)_{\bullet,v}$ is a well-defined $n \times 1$-matrix (since $v \in \{1, 2, \ldots, n\}$).

We know that C is an $n \times n$-matrix (since $C \in \mathbb{K}^{n\times n}$), and that $(I_n)_{\bullet,v}$ is an $n \times 1$-matrix. Thus, $\left(C \mid (I_n)_{\bullet,v}\right)$ is an $n \times (n+1)$-matrix. In other words, $\left(C \mid (I_n)_{\bullet,v}\right) \in \mathbb{K}^{n\times(n+1)}$.

Set $B = \left(C \mid (I_n)_{\bullet,v}\right)$. Thus, $B = \left(C \mid (I_n)_{\bullet,v}\right) \in \mathbb{K}^{n\times(n+1)}$. We have

$$B_{\bullet,r} = C_{\bullet,r} \qquad \text{for every } r \in \{1, 2, \ldots, n\} \tag{1309}$$

[581]. Furthermore,

$$\det (A \mid B_{\bullet,n+1}) = (-1)^{n+v} \det (A_{\sim v,\bullet}) \tag{1310}$$

[582]. Moreover,

$$\det (B_{\bullet,\sim r}) = (-1)^{n+v} \det (C_{\sim v,\sim r}) \qquad \text{for every } r \in \{1, 2, \ldots, n\} \tag{1311}$$

[583]. Finally,

$$B_{\bullet,\sim(n+1)} = C \tag{1313}$$

---

[581] *Proof of (1309):* Let $r \in \{1, 2, \ldots, n\}$. Proposition 6.134 **(a)** (applied to n, C, $(I_n)_{\bullet,v}$ and r instead of m, A, v and q) yields $\left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,r} = C_{\bullet,r}$. Now, from $B = \left(C \mid (I_n)_{\bullet,v}\right)$, we obtain $B_{\bullet,r} = \left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,r} = C_{\bullet,r}$. This proves (1309).

[582] *Proof of (1310):* Proposition 6.134 **(b)** (applied to n, C and $(I_n)_{\bullet,v}$ instead of m, A and v) yields $\left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,n+1} = (I_n)_{\bullet,v}$. Now, from $B = \left(C \mid (I_n)_{\bullet,v}\right)$, we obtain $B_{\bullet,n+1} = \left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,n+1} = (I_n)_{\bullet,v}$. Hence,

$$\det \left(A \mid \underbrace{B_{\bullet,n+1}}_{=(I_n)_{\bullet,v}} \right) = \det \left(A \mid (I_n)_{\bullet,v}\right) = (-1)^{n+v} \det (A_{\sim v,\bullet})$$

(by Proposition 6.135 **(b)**, applied to $p = v$). This proves (1310).

[583] *Proof of (1311):* Let $r \in \{1, 2, \ldots, n\}$. From $B = \left(C \mid (I_n)_{\bullet,v}\right)$, we obtain $B_{\bullet,\sim r} = \left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,\sim r} = \left(C_{\bullet,\sim r} \mid (I_n)_{\bullet,v}\right)$ (by Proposition 6.134 **(c)**, applied to n, C, $(I_n)_{\bullet,v}$ and r instead of m, A, v and q).

[584].

Hence,

$$\det\left(\underbrace{B_{\bullet,\sim r}}_{=\left(C_{\bullet,\sim r}\mid (I_n)_{\bullet,v}\right)}\right) = \det\left(C_{\bullet,\sim r} \mid (I_n)_{\bullet,v}\right) = (-1)^{n+v} \det\left(\left(C_{\bullet,\sim r}\right)_{\sim v,\bullet}\right) \tag{1312}$$

(by Proposition 6.135 **(b)**, applied to v and $C_{\bullet,\sim r}$ instead of p and A). But Proposition 6.130 **(c)** (applied to n, C, v and r instead of m, A, u and v) yields $\left(C_{\bullet,\sim r}\right)_{\sim v,\bullet} = \left(C_{\sim v,\bullet}\right)_{\bullet,\sim r} = C_{\sim v,\sim r}$. Now, (1312) becomes

$$\det\left(B_{\bullet,\sim r}\right) = (-1)^{n+v} \det\left(\underbrace{\left(C_{\bullet,\sim r}\right)_{\sim v,\bullet}}_{=C_{\sim v,\sim r}}\right) = (-1)^{n+v} \det\left(C_{\sim v,\sim r}\right).$$

This proves (1311).

[584]*Proof of (1313):* Proposition 6.134 **(d)** (applied to n, C and $(I_n)_{\bullet,v}$ instead of m, A and v) yields $\left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,\sim(n+1)} = C$. Now, from $B = \left(C \mid (I_n)_{\bullet,v}\right)$, we obtain $B_{\bullet,\sim(n+1)} = \left(C \mid (I_n)_{\bullet,v}\right)_{\bullet,\sim(n+1)} = C$. This proves (1313).

Now, Theorem 6.150 yields

$$\begin{aligned}
0 &= \sum_{r=1}^{n+1} (-1)^r \det\left(A \mid B_{\bullet,r}\right) \det\left(B_{\bullet,\sim r}\right) \\
&= \sum_{r=1}^{n} (-1)^r \det\left(A \mid \underbrace{B_{\bullet,r}}_{\substack{=C_{\bullet,r} \\ \text{(by (1309))}}}\right) \underbrace{\det\left(B_{\bullet,\sim r}\right)}_{\substack{=(-1)^{n+v}\det(C_{\sim v,\sim r}) \\ \text{(by (1311))}}} \\
&\qquad + (-1)^{n+1} \underbrace{\det\left(A \mid B_{\bullet,n+1}\right)}_{\substack{=(-1)^{n+v}\det(A_{\sim v,\bullet}) \\ \text{(by (1310))}}} \det\left(\underbrace{B_{\bullet,\sim(n+1)}}_{\substack{=C \\ \text{(by (1313))}}}\right) \\
&\qquad \left(\text{here, we have split off the addend for } r = n+1 \text{ from the sum}\right) \\
&= \sum_{r=1}^{n} (-1)^r \underbrace{\det\left(A \mid C_{\bullet,r}\right)(-1)^{n+v}}_{=(-1)^{n+v}\det(A \mid C_{\bullet,r})} \det\left(C_{\sim v,\sim r}\right) \\
&\qquad + \underbrace{(-1)^{n+1}(-1)^{n+v}}_{\substack{=(-1)^{(n+1)+(n+v)}=(-1)^{v+1} \\ \text{(since } (n+1)+(n+v)=2n+(v+1)\equiv v+1 \bmod 2)}} \det\left(A_{\sim v,\bullet}\right)\det C \\
&= \sum_{r=1}^{n} \underbrace{(-1)^r(-1)^{n+v}}_{=(-1)^{r+(n+v)}} \det\left(A \mid C_{\bullet,r}\right)\det\left(C_{\sim v,\sim r}\right) + \underbrace{(-1)^{v+1}}_{=-(-1)^v} \det\left(A_{\sim v,\bullet}\right)\det C \\
&= \sum_{r=1}^{n} (-1)^{r+(n+v)} \det\left(A \mid C_{\bullet,r}\right)\det\left(C_{\sim v,\sim r}\right) - (-1)^v \det\left(A_{\sim v,\bullet}\right)\det C.
\end{aligned}$$

Adding $(-1)^v \det\left(A_{\sim v,\bullet}\right)\det C$ to both sides of this equality, we obtain

$$(-1)^v \det\left(A_{\sim v,\bullet}\right)\det C = \sum_{r=1}^{n} (-1)^{r+(n+v)} \det\left(A \mid C_{\bullet,r}\right)\det\left(C_{\sim v,\sim r}\right).$$

Multiplying both sides of this equality by $(-1)^v$, we find

$$\begin{aligned}
&(-1)^v (-1)^v \det(A_{\sim v,\bullet}) \det C \\
&= (-1)^v \sum_{r=1}^{n} (-1)^{r+(n+v)} \det(A \mid C_{\bullet,r}) \det(C_{\sim v,\sim r}) \\
&= \sum_{r=1}^{n} \underbrace{(-1)^v (-1)^{r+(n+v)}}_{\substack{=(-1)^{v+(r+(n+v))}=(-1)^{n+r} \\ (\text{since } v+(r+(n+v))=2v+n+r\equiv n+r \bmod 2)}} \det(A \mid C_{\bullet,r}) \det(C_{\sim v,\sim r}) \\
&= \sum_{r=1}^{n} (-1)^{n+r} \det(A \mid C_{\bullet,r}) \det(C_{\sim v,\sim r}) \\
&= \sum_{q=1}^{n} (-1)^{n+q} \det\left(A \mid C_{\bullet,q}\right) \det\left(C_{\sim v,\sim q}\right)
\end{aligned}$$

(here, we have renamed the summation index r as q). Hence,

$$\begin{aligned}
&\sum_{q=1}^{n} (-1)^{n+q} \det\left(A \mid C_{\bullet,q}\right) \det\left(C_{\sim v,\sim q}\right) \\
&= \underbrace{(-1)^v (-1)^v}_{\substack{=(-1)^{v+v}=1 \\ (\text{since } v+v=2v \text{ is even})}} \det(A_{\sim v,\bullet}) \det C = \det(A_{\sim v,\bullet}) \det C.
\end{aligned}$$

This proves Proposition 6.137 again. □

Thus, Exercise 6.43 is solved.

7.112. Solution to Exercise 6.44

Throughout this section, we shall use the notations introduced in Definition 6.78 and in Definition 6.153.

Let us prepare for the proof of Lemma 6.158 by showing a particular case:

Lemma 7.212. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be two $n \times n$-matrices. Let Q be a subset of $\{1, 2, \ldots, n\}$. Let $k = |Q|$. Then,

$$\begin{aligned}
&\sum_{\substack{\sigma\in S_n; \\ \sigma(\{1,2,\ldots,k\})=Q}} (-1)^\sigma \left(\prod_{i\in\{1,2,\ldots,k\}} a_{i,\sigma(i)}\right) \left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{i,\sigma(i)}\right) \\
&= (-1)^{(1+2+\cdots+k)+\sum Q} \det\left(\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} B\right).
\end{aligned}$$

Proof of Lemma 7.212. We begin with some simple observations.

The definition of $\widetilde{Q}$ yields $\widetilde{Q} = \{1, 2, \ldots, n\} \setminus Q$. Since Q is a subset of $\{1, 2, \ldots, n\}$, this leads to $\left|\widetilde{Q}\right| = \underbrace{|\{1, 2, \ldots, n\}|}_{=n} - \underbrace{|Q|}_{=k} = n - k$. Thus, $n - k = \left|\widetilde{Q}\right| \geq 0$, so that $n \geq k$ and thus $k \in \{0, 1, \ldots, n\}$.

Let $(q_1, q_2, \ldots, q_k)$ be the list of all elements of Q in increasing order (with no repetitions). (This is well-defined (by Definition 2.50), because $|Q| = k$.)

Let $(r_1, r_2, \ldots, r_{n-k})$ be the list of all elements of $\widetilde{Q}$ in increasing order (with no repetitions). (This is well-defined, because $\left|\widetilde{Q}\right| = n - k$.)

The lists $w(Q)$ and $(q_1, q_2, \ldots, q_k)$ must be identical (since they are both defined to be the list of all elements of Q in increasing order (with no repetitions)). In other words, we have $w(Q) = (q_1, q_2, \ldots, q_k)$. Similarly, $w\left(\widetilde{Q}\right) = (r_1, r_2, \ldots, r_{n-k})$.

We know that $(r_1, r_2, \ldots, r_{n-k})$ is the list of all elements of $\widetilde{Q}$ in increasing order (with no repetitions). In other words, $(r_1, r_2, \ldots, r_{n-k})$ is the list of all elements of $\{1, 2, \ldots, n\} \setminus Q$ in increasing order (with no repetitions) (since $\widetilde{Q} = \{1, 2, \ldots, n\} \setminus Q$). Thus, for every $\alpha \in S_k$ and $\beta \in S_{n-k}$, we can define an element $\sigma_{Q,\alpha,\beta} \in S_n$ according to Exercise 5.14 **(a)** (applied to Q, $(q_1, q_2, \ldots, q_k)$ and $(r_1, r_2, \ldots, r_{n-k})$ instead of I, $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_{n-k})$). Consider this $\sigma_{Q,\alpha,\beta}$. Exercise 5.14 **(b)** (applied to Q, $(q_1, q_2, \ldots, q_k)$ and $(r_1, r_2, \ldots, r_{n-k})$ instead of I, $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_{n-k})$) shows that for every $\alpha \in S_k$ and $\beta \in S_{n-k}$, we have

$$\ell\left(\sigma_{Q,\alpha,\beta}\right) = \ell(\alpha) + \ell(\beta) + \sum Q - (1 + 2 + \cdots + k)$$

and

$$(-1)^{\sigma_{Q,\alpha,\beta}} = (-1)^{\alpha} \cdot (-1)^{\beta} \cdot (-1)^{\sum Q - (1+2+\cdots+k)}. \tag{1314}$$

Exercise 5.14 **(c)** (applied to Q, $(q_1, q_2, \ldots, q_k)$ and $(r_1, r_2, \ldots, r_{n-k})$ instead of I, $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_{n-k})$) shows that the map

$$S_k \times S_{n-k} \to \{\tau \in S_n \mid \tau(\{1, 2, \ldots, k\}) = Q\},$$
$$(\alpha, \beta) \mapsto \sigma_{Q,\alpha,\beta}$$

is well-defined and a bijection.

We have $w(Q) = (q_1, q_2, \ldots, q_k)$. Thus,

$$\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,k)}^{(q_1,q_2,\ldots,q_k)} A = \operatorname{sub}_{1,2,\ldots,k}^{q_1,q_2,\ldots,q_k} A = \left(a_{x,q_y}\right)_{1\leq x\leq k,\ 1\leq y\leq k}$$
$$\left(\begin{array}{c}\text{by the definition of } \operatorname{sub}_{1,2,\ldots,k}^{q_1,q_2,\ldots,q_k} A, \\ \text{since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\end{array}\right)$$
$$= \left(a_{i,q_j}\right)_{1\leq i\leq k,\ 1\leq j\leq k}$$

(here, we have renamed the index (x, y) as (i, j)). Thus,

$$\det\left(\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} A\right) = \sum_{\sigma\in S_k} (-1)^{\sigma} \prod_{i=1}^{k} a_{i,q_{\sigma(i)}}$$
$$\left(\begin{array}{c}\text{by (341), applied to } k,\ \operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} A \text{ and } a_{i,q_j} \\ \text{instead of } n,\ A \text{ and } a_{i,j}\end{array}\right)$$
$$= \sum_{\alpha\in S_k} (-1)^{\alpha} \prod_{i=1}^{k} a_{i,q_{\alpha(i)}} \qquad (1315)$$

(here, we have renamed the summation index σ as α).

Also, $w\left(\widetilde{Q}\right) = (r_1, r_2, \ldots, r_{n-k})$. Thus,

$$\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} B = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{(r_1,r_2,\ldots,r_{n-k})} B = \operatorname{sub}_{k+1,k+2,\ldots,n}^{r_1,r_2,\ldots,r_{n-k}} B = \left(b_{k+x,r_y}\right)_{1\le x\le n-k,\ 1\le y\le n-k}$$
$$\left(\begin{array}{c}\text{by the definition of } \operatorname{sub}_{k+1,k+2,\ldots,n}^{r_1,r_2,\ldots,r_{n-k}} B, \\ \text{since } B = \left(b_{i,j}\right)_{1\le i\le n,\ 1\le j\le n}\end{array}\right)$$
$$= \left(b_{k+i,r_j}\right)_{1\le i\le n-k,\ 1\le j\le n-k}$$

(here, we have renamed the index (x, y) as (i, j)). Thus,

$$\det\left(\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} B\right) = \sum_{\sigma\in S_{n-k}} (-1)^{\sigma} \prod_{i=1}^{n-k} b_{k+i,r_{\sigma(i)}}$$
$$\left(\begin{array}{c}\text{by (341), applied to } n-k,\ \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} B \text{ and } b_{k+i,r_j} \\ \text{instead of } n,\ A \text{ and } a_{i,j}\end{array}\right)$$
$$= \sum_{\beta\in S_{n-k}} (-1)^{\beta} \prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}} \qquad (1316)$$

(here, we have renamed the summation index σ as β).

Now, we claim the following: For any $\alpha \in S_k$ and $\beta \in S_{n-k}$, we have

$$\prod_{i\in\{1,2,\ldots,k\}} a_{i,\sigma_{Q,\alpha,\beta}(i)} = \prod_{i=1}^{k} a_{i,q_{\alpha(i)}} \qquad (1317)$$

and

$$\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{i,\sigma_{Q,\alpha,\beta}(i)} = \prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}}. \qquad (1318)$$

[*Proof of (1317) and (1318):* Let $\alpha \in S_k$ and $\beta \in S_{n-k}$. The permutation $\sigma_{Q,\alpha,\beta}$ was defined as the unique $\sigma \in S_n$ satisfying

$$\left(\sigma\left(1\right), \sigma\left(2\right), \ldots, \sigma\left(n\right)\right) = \left(q_{\alpha(1)}, q_{\alpha(2)}, \ldots, q_{\alpha(k)}, r_{\beta(1)}, r_{\beta(2)}, \ldots, r_{\beta(n-k)}\right). \qquad (1319)$$

Hence, $\sigma_{Q,\alpha,\beta}$ is a $\sigma \in S_n$ satisfying (1319). In other words, $\sigma_{Q,\alpha,\beta}$ is an element of S_n and satisfies

$$\left(\sigma_{Q,\alpha,\beta}\left(1\right), \sigma_{Q,\alpha,\beta}\left(2\right), \ldots, \sigma_{Q,\alpha,\beta}\left(n\right)\right) = \left(q_{\alpha(1)}, q_{\alpha(2)}, \ldots, q_{\alpha(k)}, r_{\beta(1)}, r_{\beta(2)}, \ldots, r_{\beta(n-k)}\right).$$

In other words,

$$\left(\sigma_{Q,\alpha,\beta}\left(i\right) = q_{\alpha(i)} \qquad \text{for every } i \in \{1, 2, \ldots, k\}\right) \qquad (1320)$$

and

$$\left(\sigma_{Q,\alpha,\beta}\left(i\right) = r_{\beta(i-k)} \qquad \text{for every } i \in \{k+1, k+2, \ldots, n\}\right). \qquad (1321)$$

Now,

$$\underbrace{\prod_{i \in \{1,2,\ldots,k\}}}_{=\prod_{i=1}^{k}} \underbrace{a_{i,\sigma_{Q,\alpha,\beta}(i)}}_{\substack{=a_{i,q_{\alpha(i)}} \\ \text{(since } \sigma_{Q,\alpha,\beta}(i)=q_{\alpha(i)} \\ \text{(by (1320)))}}} = \prod_{i=1}^{k} a_{i,q_{\alpha(i)}}.$$

This proves (1317). Furthermore,

$$\underbrace{\prod_{i \in \{k+1,k+2,\ldots,n\}}}_{=\prod_{i=k+1}^{n}} \underbrace{b_{i,\sigma_{Q,\alpha,\beta}(i)}}_{\substack{=b_{i,r_{\beta(i-k)}} \\ \text{(since } \sigma_{Q,\alpha,\beta}(i)=r_{\beta(i-k)} \\ \text{(by (1321)))}}} = \prod_{i=k+1}^{n} b_{i,r_{\beta(i-k)}} = \prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}}$$

(here, we have substituted $k+i$ for i in the product). This proves (1318).]

Now,

$$\underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(\{1,2,\ldots,k\})=Q}}}_{=\sum_{\sigma\in\{\tau\in S_n\ \mid\ \tau(\{1,2,\ldots,k\})=Q\}}} (-1)^{\sigma}\left(\prod_{i\in\{1,2,\ldots,k\}} a_{i,\sigma(i)}\right)\left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{i,\sigma(i)}\right)$$

$$= \sum_{\sigma\in\{\tau\in S_n\ \mid\ \tau(\{1,2,\ldots,k\})=Q\}} (-1)^{\sigma}\left(\prod_{i\in\{1,2,\ldots,k\}} a_{i,\sigma(i)}\right)\left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{i,\sigma(i)}\right)$$

$$= \underbrace{\sum_{(\alpha,\beta)\in S_k\times S_{n-k}}}_{=\sum_{\alpha\in S_k}\sum_{\beta\in S_{n-k}}} (-1)^{\sigma_{Q,\alpha,\beta}} \underbrace{\left(\prod_{i\in\{1,2,\ldots,k\}} a_{i,\sigma_{Q,\alpha,\beta}(i)}\right)}_{\substack{=\prod_{i=1}^{k} a_{i,q_{\alpha(i)}}\\ \text{(by (1317))}}} \underbrace{\left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{i,\sigma_{Q,\alpha,\beta}(i)}\right)}_{\substack{=\prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}}\\ \text{(by (1318))}}}$$

$$\left(\begin{array}{c}\text{here, we have substituted } \sigma_{Q,\alpha,\beta} \text{ for } \sigma \text{ in the sum, since the}\\ \text{map } S_k\times S_{n-k}\to\{\tau\in S_n\ \mid\ \tau(\{1,2,\ldots,k\})=Q\},\ (\alpha,\beta)\mapsto\sigma_{Q,\alpha,\beta}\\ \text{is a bijection}\end{array}\right)$$

$$= \sum_{\alpha\in S_k}\sum_{\beta\in S_{n-k}} \underbrace{(-1)^{\sigma_{Q,\alpha,\beta}}}_{\substack{=(-1)^{\alpha}\cdot(-1)^{\beta}\cdot(-1)^{\sum Q-(1+2+\cdots+k)}\\ \text{(by (1314))}}} \left(\prod_{i=1}^{k} a_{i,q_{\alpha(i)}}\right)\left(\prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}}\right)$$

$$= \sum_{\alpha\in S_k}\sum_{\beta\in S_{n-k}} (-1)^{\alpha}\cdot(-1)^{\beta}\cdot(-1)^{\sum Q-(1+2+\cdots+k)} \left(\prod_{i=1}^{k} a_{i,q_{\alpha(i)}}\right)\left(\prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}}\right)$$

$$= \underbrace{(-1)^{\sum Q-(1+2+\cdots+k)}}_{\substack{=(-1)^{(1+2+\cdots+k)+\sum Q}\\ \text{(since } \sum Q-(1+2+\cdots+k)\\ \equiv\sum Q+(1+2+\cdots+k)\\ =(1+2+\cdots+k)+\sum Q \bmod 2)}} \underbrace{\left(\sum_{\alpha\in S_k}(-1)^{\alpha}\prod_{i=1}^{k} a_{i,q_{\alpha(i)}}\right)}_{\substack{=\det\left(\operatorname{sub}^{w(Q)}_{(1,2,\ldots,k)} A\right)\\ \text{(by (1315))}}} \underbrace{\left(\sum_{\beta\in S_{n-k}}(-1)^{\beta}\prod_{i=1}^{n-k} b_{k+i,r_{\beta(i)}}\right)}_{\substack{=\det\left(\operatorname{sub}^{w(\widetilde{Q})}_{(k+1,k+2,\ldots,n)} B\right)\\ \text{(by (1316))}}}$$

$$= (-1)^{(1+2+\cdots+k)+\sum Q}\det\left(\operatorname{sub}^{w(Q)}_{(1,2,\ldots,k)} A\right)\det\left(\operatorname{sub}^{w(\widetilde{Q})}_{(k+1,k+2,\ldots,n)} B\right).$$

This proves Lemma 7.212. □

Before we move on to the proof of Lemma 6.158, let us show two more lemmas. The first one is a really simple fact about symmetric groups:

Lemma 7.213. Let $n \in \mathbb{N}$. Let $\gamma \in S_n$. Then, the map $S_n \to S_n$, $\sigma \mapsto \sigma \circ \gamma$ is a bijection.

Proof of Lemma 7.213. It is easy to see that the maps $S_n \to S_n$, $\sigma \mapsto \sigma \circ \gamma$ and $S_n \to S_n$, $\sigma \mapsto \sigma \circ \gamma^{-1}$ are mutually inverse. Thus, the map $S_n \to S_n$, $\sigma \mapsto \sigma \circ \gamma$ is invertible, i.e., a bijection. Lemma 7.213 is proven. □

Next, we show a lemma which is a distillate of some parts of Exercise 5.14:

Lemma 7.214. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let I be a subset of $\{1, 2, \ldots, n\}$. Let $k = |I|$. Then, there exists a $\sigma \in S_n$ satisfying $(\sigma(1), \sigma(2), \ldots, \sigma(k)) = w(I)$, $(\sigma(k+1), \sigma(k+2), \ldots, \sigma(n)) = w\left(\widetilde{I}\right)$ and $(-1)^{\sigma} = (-1)^{\sum I - (1+2+\cdots+k)}$.

Proof of Lemma 7.214. We begin by introducing some notations:

The definition of $\widetilde{I}$ yields $\widetilde{I} = \{1, 2, \ldots, n\} \setminus I$. Since I is a subset of $\{1, 2, \ldots, n\}$, this leads to $\left|\widetilde{I}\right| = \underbrace{|\{1, 2, \ldots, n\}|}_{=n} - \underbrace{|I|}_{=k} = n - k$. Thus, $n - k = \left|\widetilde{I}\right| \geq 0$, so that $n \geq k$ and thus $k \in \{0, 1, \ldots, n\}$.

We know that $w(I)$ is the list of all elements of I in increasing order (with no repetitions) (by the definition of $w(I)$). Thus, $w(I)$ is a list of $|I|$ elements. In other words, $w(I)$ is a list of k elements (since $|I| = k$).

Write $w(I)$ in the form $w(I) = (a_1, a_2, \ldots, a_k)$. (This is possible, since $w(I)$ is a list of k elements.)

We know that $w(I)$ is the list of all elements of I in increasing order (with no repetitions). In other words, $(a_1, a_2, \ldots, a_k)$ is the list of all elements of I in increasing order (with no repetitions) (since $w(I) = (a_1, a_2, \ldots, a_k)$).

We know that $w\left(\widetilde{I}\right)$ is the list of all elements of $\widetilde{I}$ in increasing order (with no repetitions) (by the definition of $w\left(\widetilde{I}\right)$). Thus, $w\left(\widetilde{I}\right)$ is a list of $\left|\widetilde{I}\right|$ elements. In other words, $w\left(\widetilde{I}\right)$ is a list of $n - k$ elements (since $\left|\widetilde{I}\right| = n - k$).

Write $w\left(\widetilde{I}\right)$ in the form $w\left(\widetilde{I}\right) = (b_1, b_2, \ldots, b_{n-k})$. (This is possible, since $w\left(\widetilde{I}\right)$ is a list of $n - k$ elements.)

We know that $w\left(\widetilde{I}\right)$ is the list of all elements of $\widetilde{I}$ in increasing order (with no repetitions). In other words, $(b_1, b_2, \ldots, b_{n-k})$ is the list of all elements of $\widetilde{I}$ in increasing order (with no repetitions) (since $w\left(\widetilde{I}\right) = (b_1, b_2, \ldots, b_{n-k})$). In other words, $(b_1, b_2, \ldots, b_{n-k})$ is the list of all elements of $\{1, 2, \ldots, n\} \setminus I$ in increasing order (with no repetitions) (since $\widetilde{I} = \{1, 2, \ldots, n\} \setminus I$).

Now we know that $(a_1, a_2, \ldots, a_k)$ is the list of all elements of I in increasing order (with no repetitions), and that $(b_1, b_2, \ldots, b_{n-k})$ is the list of all elements of

$\{1,2,\ldots,n\} \setminus I$ in increasing order (with no repetitions). Hence, for every $\alpha \in S_k$ and $\beta \in S_{n-k}$, we can define an element $\sigma_{I,\alpha,\beta} \in S_n$ according to Exercise 5.14 **(a)**. Consider this $\sigma_{I,\alpha,\beta}$. Exercise 5.14 **(b)** shows that for every $\alpha \in S_k$ and $\beta \in S_{n-k}$, we have

$$\ell\left(\sigma_{I,\alpha,\beta}\right) = \ell\left(\alpha\right) + \ell\left(\beta\right) + \sum I - \left(1+2+\cdots+k\right)$$

and

$$(-1)^{\sigma_{I,\alpha,\beta}} = (-1)^{\alpha} \cdot (-1)^{\beta} \cdot (-1)^{\sum I - (1+2+\cdots+k)}. \tag{1322}$$

Now, consider the identity permutations $\operatorname{id} \in S_k$ and $\operatorname{id} \in S_{n-k}$. (These are (in general) two different permutations, although we denote them both by id.) They give rise to an element $\sigma_{I,\operatorname{id},\operatorname{id}} \in S_n$ (obtained by setting $\alpha = \operatorname{id} \in S_k$ and $\beta = \operatorname{id} \in S_{n-k}$ in the definition of $\sigma_{I,\alpha,\beta}$). Denote this element $\sigma_{I,\operatorname{id},\operatorname{id}}$ by γ. Thus, $\gamma = \sigma_{I,\operatorname{id},\operatorname{id}}$, so that

$$\begin{aligned}(-1)^{\gamma} &= (-1)^{\sigma_{I,\operatorname{id},\operatorname{id}}} = \underbrace{(-1)^{\operatorname{id}}}_{=1} \cdot \underbrace{(-1)^{\operatorname{id}}}_{=1} \cdot (-1)^{\sum I - (1+2+\cdots+k)} \\ &\qquad \left(\text{by (1322) (applied to } \alpha = \operatorname{id} \text{ and } \beta = \operatorname{id}\text{)}\right) \\ &= (-1)^{\sum I - (1+2+\cdots+k)}. \end{aligned} \tag{1323}$$

Now, we claim that

$$\left(\gamma\left(1\right), \gamma\left(2\right), \ldots, \gamma\left(k\right)\right) = \left(a_1, a_2, \ldots, a_k\right) \tag{1324}$$

and

$$\left(\gamma\left(k+1\right), \gamma\left(k+2\right), \ldots, \gamma\left(n\right)\right) = \left(b_1, b_2, \ldots, b_{n-k}\right). \tag{1325}$$

[*Proof of (1324) and (1325):* The permutation γ is $\sigma_{I,\operatorname{id},\operatorname{id}}$. In other words, the permutation γ is the unique $\sigma \in S_n$ satisfying

$$\left(\sigma\left(1\right), \sigma\left(2\right), \ldots, \sigma\left(n\right)\right) = \left(a_{\operatorname{id}(1)}, a_{\operatorname{id}(2)}, \ldots, a_{\operatorname{id}(k)}, b_{\operatorname{id}(1)}, b_{\operatorname{id}(2)}, \ldots, b_{\operatorname{id}(n-k)}\right) \tag{1326}$$

(because this is how $\sigma_{I,\operatorname{id},\operatorname{id}}$ was defined). Hence, γ is a $\sigma \in S_n$ satisfying (1326). In other words, γ is an element of S_n and satisfies

$$\left(\gamma\left(1\right), \gamma\left(2\right), \ldots, \gamma\left(n\right)\right) = \left(a_{\operatorname{id}(1)}, a_{\operatorname{id}(2)}, \ldots, a_{\operatorname{id}(k)}, b_{\operatorname{id}(1)}, b_{\operatorname{id}(2)}, \ldots, b_{\operatorname{id}(n-k)}\right).$$

Thus,

$$\begin{aligned}\left(\gamma\left(1\right), \gamma\left(2\right), \ldots, \gamma\left(n\right)\right) &= \left(a_{\operatorname{id}(1)}, a_{\operatorname{id}(2)}, \ldots, a_{\operatorname{id}(k)}, b_{\operatorname{id}(1)}, b_{\operatorname{id}(2)}, \ldots, b_{\operatorname{id}(n-k)}\right) \\ &= \left(a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_{n-k}\right).\end{aligned}$$

Hence,

$$\left(\gamma\left(i\right) = a_i \qquad \text{for every } i \in \{1,2,\ldots,k\}\right) \tag{1327}$$

and

$$\left(\gamma\left(i\right) = b_{i-k} \qquad \text{for every } i \in \{k+1,k+2,\ldots,n\}\right). \tag{1328}$$

Now, (1324) follows immediately from (1327). Furthermore, (1325) follows immediately from (1328). Hence, both (1324) and (1325) are proven.]

Comparing (1324) with $w(I) = (a_1, a_2, \ldots, a_k)$, we obtain

$$(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(I).$$

Comparing (1325) with $w\left(\widetilde{I}\right) = (b_1, b_2, \ldots, b_{n-k})$, we obtain

$$(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{I}\right).$$

Now, we have shown that the permutation $\gamma \in S_n$ satisfies $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(I)$, $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{I}\right)$ and $(-1)^\gamma = (-1)^{\sum I-(1+2+\cdots+k)}$. Hence, there exists a $\sigma \in S_n$ satisfying $(\sigma(1), \sigma(2), \ldots, \sigma(k)) = w(I)$, $(\sigma(k+1), \sigma(k+2), \ldots, \sigma(n)) = w\left(\widetilde{I}\right)$ and $(-1)^\sigma = (-1)^{\sum I-(1+2+\cdots+k)}$ (namely, $\sigma = \gamma$). This proves Lemma 7.214. $\square$

Proof of Lemma 6.158. Let us first study the sets P and $\widetilde{P}$.

Define $k \in \mathbb{N}$ by $k = |P| = |Q|$. (This makes sense, since we assumed that $|P| = |Q|$.)

The definition of $\widetilde{P}$ yields $\widetilde{P} = \{1, 2, \ldots, n\} \setminus P$. Since P is a subset of $\{1, 2, \ldots, n\}$, this leads to $\left|\widetilde{P}\right| = \underbrace{|\{1, 2, \ldots, n\}|}_{=n} - \underbrace{|P|}_{=k} = n - k$. Thus, $n - k = \left|\widetilde{P}\right| \geq 0$, so that $n \geq k$ and thus $k \in \{0, 1, \ldots, n\}$.

Let $(p_1, p_2, \ldots, p_k)$ be the list of all elements of P in increasing order (with no repetitions). (This is well-defined, because $|P| = k$.)

Let $(r_1, r_2, \ldots, r_{n-k})$ be the list of all elements of $\widetilde{P}$ in increasing order (with no repetitions). (This is well-defined, because $\left|\widetilde{P}\right| = n - k$.)

Lemma 7.214 (applied to $I = P$) shows that there exists a $\sigma \in S_n$ satisfying $(\sigma(1), \sigma(2), \ldots, \sigma(k)) = w(P)$, $(\sigma(k+1), \sigma(k+2), \ldots, \sigma(n)) = w\left(\widetilde{P}\right)$ and $(-1)^\sigma = (-1)^{\sum P-(1+2+\cdots+k)}$. Denote this σ by γ. Thus, γ is an element of S_n satisfying $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(P)$, $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{P}\right)$ and $(-1)^\gamma = (-1)^{\sum P-(1+2+\cdots+k)}$.

Notice that

$$\begin{aligned}(-1)^{\sum P-(1+2+\cdots+k)} \underbrace{(-1)^\gamma}_{=(-1)^{\sum P-(1+2+\cdots+k)}} &= (-1)^{\sum P-(1+2+\cdots+k)} (-1)^{\sum P-(1+2+\cdots+k)} \\ &= \left((-1)^{\sum P-(1+2+\cdots+k)}\right)^2 = (-1)^{2(\sum P-(1+2+\cdots+k))} \\ &= 1 \qquad (1329)\end{aligned}$$

(since $2\left(\sum P - (1+2+\cdots+k)\right)$ is even).

Now, we claim that

$$(\gamma(i) = p_i \qquad \text{for every } i \in \{1, 2, \ldots, k\}) \tag{1330}$$

and

$$(\gamma(k+i) = r_i \qquad \text{for every } i \in \{1, 2, \ldots, n-k\}). \tag{1331}$$

[*Proof of (1330) and (1331):* The lists $w(P)$ and $(p_1, p_2, \ldots, p_k)$ must be identical (since they are both defined to be the list of all elements of P in increasing order (with no repetitions)). In other words, we have $w(P) = (p_1, p_2, \ldots, p_k)$. Now,

$$(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(P) = (p_1, p_2, \ldots, p_k).$$

In other words, $\gamma(i) = p_i$ for every $i \in \{1, 2, \ldots, k\}$. This proves (1330).

The lists $w\left(\widetilde{P}\right)$ and $(r_1, r_2, \ldots, r_{n-k})$ must be identical (since they are both defined to be the list of all elements of $\widetilde{P}$ in increasing order (with no repetitions)). In other words, we have $w\left(\widetilde{P}\right) = (r_1, r_2, \ldots, r_{n-k})$. Now,

$$(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{P}\right) = (r_1, r_2, \ldots, r_{n-k}).$$

In other words, $\gamma(k+i) = r_i$ for every $i \in \{1, 2, \ldots, n-k\}$. This proves (1331). Thus, both (1330) and (1331) are proven.]

Define an $n \times n$-matrix A' by $A' = \left(a_{\gamma(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Define an $n \times n$-matrix B' by $B' = \left(b_{\gamma(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Then,

$$\operatorname{sub}_{w(P)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} \left(A'\right) \tag{1332}$$

[585] and

$$\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} \left(B'\right) \tag{1334}$$

[585] *Proof of (1332):* Write the list $w(Q)$ in the form $w(Q) = (q_1, q_2, \ldots, q_\ell)$ for some $\ell \in \mathbb{N}$. (Actually, $\ell = k$, but we will not use this.)

Recall that $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(P)$, so that $w(P) = (\gamma(1), \gamma(2), \ldots, \gamma(k))$.

From $w(P) = (\gamma(1), \gamma(2), \ldots, \gamma(k))$ and $w(Q) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{w(P)}^{w(Q)} A = \operatorname{sub}_{(\gamma(1),\gamma(2),\ldots,\gamma(k))}^{(q_1,q_2,\ldots,q_\ell)} A = \operatorname{sub}_{\gamma(1),\gamma(2),\ldots,\gamma(k)}^{q_1,q_2,\ldots,q_\ell} A = \left(a_{\gamma(x),q_y}\right)_{1 \leq x \leq k,\ 1 \leq y \leq \ell} \tag{1333}$$

$$\left(\text{by the definition of } \operatorname{sub}_{\gamma(1),\gamma(2),\ldots,\gamma(k)}^{q_1,q_2,\ldots,q_\ell} A, \text{ since } A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}\right).$$

On the other hand, from $w(Q) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} \left(A'\right) = \operatorname{sub}_{(1,2,\ldots,k)}^{(q_1,q_2,\ldots,q_\ell)} \left(A'\right) = \operatorname{sub}_{1,2,\ldots,k}^{q_1,q_2,\ldots,q_\ell} \left(A'\right) = \left(a_{\gamma(x),q_y}\right)_{1 \leq x \leq k,\ 1 \leq y \leq \ell}$$

(by the definition of $\operatorname{sub}_{1,2,\ldots,k}^{q_1,q_2,\ldots,q_\ell} \left(A'\right)$, since $A' = \left(a_{\gamma(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$). Comparing this with (1333), we obtain $\operatorname{sub}_{w(P)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} \left(A'\right)$. This proves (1332).

[586].

Lemma 7.213 shows that the map $S_n \to S_n,\ \sigma \mapsto \sigma \circ \gamma$ is a bijection. But $\gamma(\{1,2,\ldots,k\}) = P$ [587]. Hence, every $\sigma \in S_n$ satisfies

$$(\sigma \circ \gamma)(\{1,2,\ldots,k\}) = \sigma\left(\underbrace{\gamma(\{1,2,\ldots,k\})}_{=P}\right) = \sigma(P). \qquad (1336)$$

Furthermore, every $\sigma \in S_n$ satisfies

$$\prod_{i\in\{1,2,\ldots,k\}} a_{\gamma(i),(\sigma\circ\gamma)(i)} = \prod_{i\in P} a_{i,\sigma(i)} \qquad (1337)$$

[588] and

$$\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{\gamma(i),(\sigma\circ\gamma)(i)} = \prod_{i\in\widetilde{P}} b_{i,\sigma(i)} \qquad (1339)$$

---

[586] *Proof of (1334):* Write the list $w\left(\widetilde{Q}\right)$ in the form $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$ for some $\ell \in \mathbb{N}$. (Actually, $\ell = n-k$, but we will not use this.)

Recall that $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{P}\right)$. Thus, $w\left(\widetilde{P}\right) = (\gamma(k+1), \gamma(k+2), \ldots, \gamma(n))$.

From $w\left(\widetilde{P}\right) = (\gamma(k+1), \gamma(k+2), \ldots, \gamma(n))$ and $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B = \operatorname{sub}_{(\gamma(k+1),\gamma(k+2),\ldots,\gamma(n))}^{(q_1,q_2,\ldots,q_\ell)} B = \operatorname{sub}_{\gamma(k+1),\gamma(k+2),\ldots,\gamma(n)}^{q_1,q_2,\ldots,q_\ell} B = \left(b_{\gamma(k+x),q_y}\right)_{1\le x\le n-k,\ 1\le y\le \ell} \qquad (1335)$$

$$\left(\text{by the definition of } \operatorname{sub}_{\gamma(k+1),\gamma(k+2),\ldots,\gamma(n)}^{q_1,q_2,\ldots,q_\ell} B, \text{ since } B = \left(b_{i,j}\right)_{1\le i\le n,\ 1\le j\le n}\right).$$

On the other hand, from $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})}(B') = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{(q_1,q_2,\ldots,q_\ell)}(B') = \operatorname{sub}_{k+1,k+2,\ldots,n}^{q_1,q_2,\ldots,q_\ell}(B') = \left(b_{\gamma(k+x),q_y}\right)_{1\le x\le n-k,\ 1\le y\le \ell}$$

(by the definition of $\operatorname{sub}_{k+1,k+2,\ldots,n}^{q_1,q_2,\ldots,q_\ell}(B')$, since $B' = \left(b_{\gamma(i),j}\right)_{1\le i\le n,\ 1\le j\le n}$). Comparing this with (1335), we obtain $\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})}(B')$. This proves (1334).

[587] *Proof.* We have

$$\gamma(\{1,2,\ldots,k\}) = \{\gamma(1), \gamma(2), \ldots, \gamma(k)\} = \left\{\underbrace{\gamma(i)}_{\substack{=p_i \\ \text{(by (1330))}}} \mid i \in \{1,2,\ldots,k\}\right\}$$
$$= \{p_i \mid i \in \{1,2,\ldots,k\}\} = \{p_1, p_2, \ldots, p_k\} = P$$

(since $(p_1, p_2, \ldots, p_k)$ is a list of all elements of P). Qed.

[588] *Proof of (1337):* Let $\sigma \in S_n$.

We shall use the notation introduced in Definition 7.84. Thus, $[k] = \{1, 2, \ldots, k\}$.

[589].

Now, Lemma 7.212 (applied to A', $a_{\gamma(i),j}$, B' and $b_{\gamma(i),j}$ instead of A, $a_{i,j}$, B and

---

Every $i \in \{1,2,\ldots,k\}$ satisfies $p_i = \gamma(i)$ (by (1330)) and $\sigma\left(\underbrace{p_i}_{=\gamma(i)}\right) = \sigma(\gamma(i)) = (\sigma\circ\gamma)(i)$.

Hence, every $i \in \{1,2,\ldots,k\}$ satisfies

$$a_{p_i,\sigma(p_i)} = a_{\gamma(i),(\sigma\circ\gamma)(i)}. \tag{1338}$$

Recall that $(p_1, p_2, \ldots, p_k)$ is the list of all elements of P in increasing order (with no repetitions). Hence, Lemma 7.89 **(a)** (applied to P, k and $(p_1, p_2, \ldots, p_k)$ instead of S, s and $(c_1, c_2, \ldots, c_s)$) shows that the map $[k] \to P,\ h \mapsto p_h$ is well-defined and a bijection. Hence, we can substitute p_h for i in the product $\prod_{i\in P} a_{i,\sigma(i)}$. We thus obtain

$$\prod_{i\in P} a_{i,\sigma(i)} = \prod_{h\in[k]} a_{p_h,\sigma(p_h)} = \underbrace{\prod_{i\in[k]}}_{\substack{=\prod_{i\in\{1,2,\ldots,k\}} \\ (\text{since } [k]=\{1,2,\ldots,k\})}} a_{p_i,\sigma(p_i)}$$

(here, we have renamed the index h as i in the product)

$$= \prod_{i\in\{1,2,\ldots,k\}} \underbrace{a_{p_i,\sigma(p_i)}}_{\substack{=a_{\gamma(i),(\sigma\circ\gamma)(i)} \\ (\text{by } (1338))}} = \prod_{i\in\{1,2,\ldots,k\}} a_{\gamma(i),(\sigma\circ\gamma)(i)}.$$

This proves (1337).

[589] *Proof of (1339):* Let $\sigma \in S_n$.

We shall use the notation introduced in Definition 7.84. Thus, $[n-k] = \{1,2,\ldots,n-k\}$.

Every $i \in \{1,2,\ldots,n-k\}$ satisfies $r_i = \gamma(k+i)$ (by (1331)) and $\sigma\left(\underbrace{r_i}_{=\gamma(k+i)}\right) = \sigma(\gamma(k+i)) = (\sigma\circ\gamma)(k+i)$. Hence, every $i \in \{1,2,\ldots,n-k\}$ satisfies

$$b_{r_i,\sigma(r_i)} = b_{\gamma(k+i),(\sigma\circ\gamma)(k+i)}. \tag{1340}$$

Recall that $(r_1, r_2, \ldots, r_{n-k})$ is the list of all elements of $\widetilde{P}$ in increasing order (with no repetitions). Hence, Lemma 7.89 **(a)** (applied to $\widetilde{P}$, $n-k$ and $(r_1, r_2, \ldots, r_{n-k})$ instead of S, s and $(c_1, c_2, \ldots, c_s)$) shows that the map $[n-k] \to \widetilde{P},\ h \mapsto r_h$ is well-defined and a bijection. Hence,

$b_{i,j}$) yields

$$\sum_{\substack{\sigma\in S_n;\\ \sigma(\{1,2,\ldots,k\})=Q}} (-1)^{\sigma} \left(\prod_{i\in\{1,2,\ldots,k\}} a_{\gamma(i),\sigma(i)}\right) \left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{\gamma(i),\sigma(i)}\right)$$
$$= (-1)^{(1+2+\cdots+k)+\sum Q} \det \underbrace{\left(\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} (A')\right)}_{\substack{=\operatorname{sub}_{w(P)}^{w(Q)} A \\ \text{(by (1332))}}} \det \underbrace{\left(\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} (B')\right)}_{\substack{=\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B \\ \text{(by (1334))}}}$$
$$= (-1)^{(1+2+\cdots+k)+\sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right).$$

---

we can substitute r_h for i in the product $\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}$. We thus obtain

$$\prod_{i\in\widetilde{P}} b_{i,\sigma(i)} = \prod_{h\in[n-k]} b_{r_h,\sigma(r_h)} = \underbrace{\prod_{i\in[n-k]}}_{\substack{=\prod_{i\in\{1,2,\ldots,n-k\}} \\ \text{(since } [n-k]=\{1,2,\ldots,n-k\})}} b_{r_i,\sigma(r_i)}$$
$$\text{(here, we have renamed the index } h \text{ as } i \text{ in the product)}$$
$$= \prod_{i\in\{1,2,\ldots,n-k\}} \underbrace{b_{r_i,\sigma(r_i)}}_{\substack{=b_{\gamma(k+i),(\sigma\circ\gamma)(k+i)} \\ \text{(by (1340))}}}$$
$$= \prod_{i\in\{1,2,\ldots,n-k\}} b_{\gamma(k+i),(\sigma\circ\gamma)(k+i)} = \prod_{i\in\{k+1,k+2,\ldots,n\}} b_{\gamma(i),(\sigma\circ\gamma)(i)}.$$
$$\text{(here, we have substituted } i \text{ for } k+i \text{ in the product)}.$$

This proves (1339).

Comparing this with

$$\begin{aligned}
&\sum_{\substack{\sigma\in S_n;\\ \sigma(\{1,2,\ldots,k\})=Q}} (-1)^{\sigma} \left(\prod_{i\in\{1,2,\ldots,k\}} a_{\gamma(i),\sigma(i)}\right)\left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{\gamma(i),\sigma(i)}\right)\\
&= \underbrace{\sum_{\substack{\sigma\in S_n;\\ (\sigma\circ\gamma)(\{1,2,\ldots,k\})=Q}}}_{\substack{=\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}}\\ \text{(since every } \sigma\in S_n\\ \text{satisfies } (\sigma\circ\gamma)(\{1,2,\ldots,k\})=\sigma(P)\\ \text{(by (1336)))}}} \underbrace{(-1)^{\sigma\circ\gamma}}_{\substack{=(-1)^{\sigma}\cdot(-1)^{\gamma}\\ \text{(by (315), applied}\\ \text{to } \tau=\gamma)}}\\
&\qquad \underbrace{\left(\prod_{i\in\{1,2,\ldots,k\}} a_{\gamma(i),(\sigma\circ\gamma)(i)}\right)}_{\substack{=\prod_{i\in P} a_{i,\sigma(i)}\\ \text{(by (1337))}}}\underbrace{\left(\prod_{i\in\{k+1,k+2,\ldots,n\}} b_{\gamma(i),(\sigma\circ\gamma)(i)}\right)}_{\substack{=\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\\ \text{(by (1339))}}}\\
&\qquad \left(\begin{array}{c}\text{here, we have substituted } \sigma\circ\gamma \text{ for } \sigma \text{ in the sum, since}\\ \text{the map } S_n\to S_n,\ \sigma\mapsto\sigma\circ\gamma \text{ is a bijection}\end{array}\right)\\
&= \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\cdot(-1)^{\gamma}\left(\prod_{i\in P} a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\right)\\
&= (-1)^{\gamma}\cdot\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\left(\prod_{i\in P} a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\right),
\end{aligned}$$

we obtain

$$\begin{aligned}
&(-1)^{(1+2+\cdots+k)+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right)\\
&= (-1)^{\gamma}\cdot\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\left(\prod_{i\in P} a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\right).
\end{aligned}$$

Multiplying both sides of this equality by $(-1)^{\sum P-(1+2+\cdots+k)}$, we obtain

$$(-1)^{\sum P-(1+2+\cdots+k)} (-1)^{(1+2+\cdots+k)+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right)$$
$$= \underbrace{(-1)^{\sum P-(1+2+\cdots+k)} (-1)^{\gamma}}_{\substack{=1 \\ \text{(by (1329))}}} \cdot \sum_{\substack{\sigma\in S_n; \\ \sigma(P)=Q}} (-1)^{\sigma} \left(\prod_{i\in P} a_{i,\sigma(i)}\right) \left(\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\right)$$
$$= \sum_{\substack{\sigma\in S_n; \\ \sigma(P)=Q}} (-1)^{\sigma} \left(\prod_{i\in P} a_{i,\sigma(i)}\right) \left(\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\right).$$

Thus,

$$\sum_{\substack{\sigma\in S_n; \\ \sigma(P)=Q}} (-1)^{\sigma} \left(\prod_{i\in P} a_{i,\sigma(i)}\right) \left(\prod_{i\in\widetilde{P}} b_{i,\sigma(i)}\right)$$
$$= \underbrace{(-1)^{\sum P-(1+2+\cdots+k)} (-1)^{(1+2+\cdots+k)+\sum Q}}_{\substack{=(-1)^{\left(\sum P-(1+2+\cdots+k)\right)+\left((1+2+\cdots+k)+\sum Q\right)} \\ =(-1)^{\sum P+\sum Q} \\ \text{(since } \left(\sum P-(1+2+\cdots+k)\right)+\left((1+2+\cdots+k)+\sum Q\right)=\sum P+\sum Q\text{)}}} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right)$$
$$= (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right).$$

This proves Lemma 6.158. □

For the sake of convenience, let us restate a simplified particular case of Lemma 6.158 for $A = B$:

Lemma 7.215. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n \times n$-matrix. Let P and Q be two subsets of $\{1, 2, \ldots, n\}$ such that $|P| = |Q|$. Then,

$$\sum_{\substack{\sigma\in S_n; \\ \sigma(P)=Q}} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} = (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A\right).$$

Proof of Lemma 7.215. Every $\sigma \in S_n$ satisfies

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = \left(\prod_{i\in P} a_{i,\sigma(i)}\right) \left(\prod_{i\in\widetilde{P}} a_{i,\sigma(i)}\right) \tag{1341}$$

[590]. Now,

$$\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{=\left(\prod_{i\in P} a_{i,\sigma(i)}\right)\left(\prod_{i\in \widetilde{P}} a_{i,\sigma(i)}\right)\\ \text{(by (1341))}}}$$
$$= \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma} \left(\prod_{i\in P} a_{i,\sigma(i)}\right)\left(\prod_{i\in \widetilde{P}} a_{i,\sigma(i)}\right)$$
$$= (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w\left(\widetilde{Q}\right)} A\right)$$

(by Lemma 6.158 (applied to $B = A$ and $b_{i,j} = a_{i,j}$)). This proves Lemma 7.215. □

Another fact that we will need is very simple:

Lemma 7.216. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$. Let P be a subset of $\{1,2,\ldots,n\}$.
(a) The set $\sigma(P)$ is a subset of $\{1,2,\ldots,n\}$ satisfying $|\sigma(P)| = |P|$.
(b) Let Q be a subset of $\{1,2,\ldots,n\}$. Then, $\sigma^{-1}(Q) = P$ holds if and only if $\sigma(P) = Q$.

Proof of Lemma 7.216. We have $\sigma \in S_n$. In other words, σ is a permutation of the set $\{1,2,\ldots,n\}$. In other words, σ is a bijective map $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$. The map σ is bijective, and therefore injective.

(a) Lemma 1.3 **(c)** (applied to $U = \{1,2,\ldots,n\}$, $V = \{1,2,\ldots,n\}$, $f = \sigma$ and $S = P$) shows that $|\sigma(P)| = |P|$ (since the map σ is injective). This proves Lemma 7.216 **(a)**.

[590]*Proof of (1341):* Let $\sigma \in S_n$. Notice that $\widetilde{P} = \{1,2,\ldots,n\} \setminus P$ (by the definition of $\widetilde{P}$). Now,

$$\underbrace{\prod_{i=1}^{n}}_{=\prod_{i\in\{1,2,\ldots,n\}}} a_{i,\sigma(i)} = \prod_{i\in\{1,2,\ldots,n\}} a_{i,\sigma(i)} = \left(\underbrace{\prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\in P}}}_{\substack{=\prod_{i\in P}\\ (\text{since } P\subseteq\{1,2,\ldots,n\})}} a_{i,\sigma(i)}\right)\left(\underbrace{\prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\notin P}}}_{\substack{=\prod_{i\in\{1,2,\ldots,n\}\setminus P}=\prod_{i\in\widetilde{P}}\\ (\text{since } \{1,2,\ldots,n\}\setminus P=\widetilde{P})}} a_{i,\sigma(i)}\right)$$
$$\left(\begin{array}{c}\text{since every } i\in\{1,2,\ldots,n\} \text{ satisfies either } i\in P\\ \text{or } i\notin P \text{ (but not both)}\end{array}\right)$$
$$= \left(\prod_{i\in P} a_{i,\sigma(i)}\right)\left(\prod_{i\in \widetilde{P}} a_{i,\sigma(i)}\right).$$

This proves (1341).

(b) We have $\sigma\left(\sigma^{-1}(Q)\right) = Q$ (since σ is bijective). But if U and V are two subsets of $\{1,2,\ldots,n\}$, then $U = V$ holds if and only if $\sigma(U) = \sigma(V)$ (because σ is bijective). Applying this to $U = \sigma^{-1}(Q)$ and $V = P$, we conclude that $\sigma^{-1}(Q) = P$ holds if and only if $\sigma\left(\sigma^{-1}(Q)\right) = \sigma(P)$. Hence, we have the following chain of equivalences:

$$\left(\sigma^{-1}(Q) = P\right) \iff \left(\underbrace{\sigma\left(\sigma^{-1}(Q)\right)}_{=Q} = \sigma(P)\right) \iff (Q = \sigma(P)) \iff (\sigma(P) = Q).$$

In other words, $\sigma^{-1}(Q) = P$ holds if and only if $\sigma(P) = Q$. Lemma 7.216 **(b)** is now proven. □

We can now step to the proof of Theorem 6.156:

Proof of Theorem 6.156. Write the $n \times n$-matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

If P and Q are two subsets of $\{1,2,\ldots,n\}$ satisfying $|Q| \neq |P|$, then

$$\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} = 0 \qquad (1342)$$

[591].

[591] *Proof of (1342):* Let P and Q be two subsets of $\{1,2,\ldots,n\}$ satisfying $|Q| \neq |P|$.

Let $\sigma \in S_n$ be such that $\sigma(P) = Q$. We shall derive a contradiction.

Indeed, Lemma 7.216 **(a)** shows that the set $\sigma(P)$ is a subset of $\{1,2,\ldots,n\}$ satisfying $|\sigma(P)| = |P|$. Hence, $|P| = \left|\underbrace{\sigma(P)}_{=Q}\right| = |Q| \neq |P|$. This is absurd. Hence, we have found a contradiction.

Now, forget that we fixed σ. We thus have found a contradiction for every $\sigma \in S_n$ satisfying $\sigma(P) = Q$. Thus, there exists no $\sigma \in S_n$ satisfying $\sigma(P) = Q$. Hence, the sum $\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}$ is an empty sum. Thus,

$$\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} = (\text{empty sum}) = 0.$$

This proves (1342).

(a) Let P be a subset of $\{1,2,\ldots,n\}$. Then, (341) yields

$$\det A = \underbrace{\sum_{\sigma\in S_n}}_{\substack{=\sum_{Q\subseteq\{1,2,\ldots,n\}}\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} \\ \text{(because for every } \sigma\in S_n\text{, the set} \\ \sigma(P)\text{ is a subset of }\{1,2,\ldots,n\})}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} = \sum_{Q\subseteq\{1,2,\ldots,n\}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} + \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|\neq|P|}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}}_{\substack{=0\\ \text{(by (1342))}}}$$

$$\left(\begin{array}{c}\text{since every subset } Q \text{ of } \{1,2,\ldots,n\} \text{ satisfies}\\ \text{either } |Q|=|P| \text{ or } |Q|\neq|P| \text{ (but not both)}\end{array}\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)} + \underbrace{\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|\neq|P|}} 0}_{=0}$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^\sigma \prod_{i=1}^n a_{i,\sigma(i)}}_{\substack{=(-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A\right) \\ \text{(by Lemma 7.215} \\ \text{(since } |P|=|Q| \text{ (since } |Q|=|P|\text{)))}}}$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A\right).$$

This proves Theorem 6.156 **(a)**.

(b) Let Q be a subset of $\{1,2,\ldots,n\}$. Then, (341) yields

$$\det A = \underbrace{\sum_{\sigma\in S_n}}_{\substack{=\sum_{P\subseteq\{1,2,\ldots,n\}}\sum_{\substack{\sigma\in S_n;\\ \sigma^{-1}(Q)=P}} \\ \text{(because for every } \sigma\in S_n\text{, the set} \\ \sigma^{-1}(Q)\text{ is a subset of }\{1,2,\ldots,n\})}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)}$$

$$= \sum_{P\subseteq\{1,2,\ldots,n\}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma^{-1}(Q)=P}}}_{\substack{=\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} \\ \text{(because for every }\sigma\in S_n\text{,} \\ \text{the statement }\left(\sigma^{-1}(Q)=P\right) \\ \text{is equivalent to} \\ \text{the statement }(\sigma(P)=Q) \\ \text{(by Lemma 7.216 \textbf{(b)}))}}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)}$$

$$= \sum_{P\subseteq\{1,2,\ldots,n\}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)}$$

$$= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)} + \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|\neq|Q|}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{=0 \\ \text{(by (1342))}}}$$

$$\left(\begin{array}{c}\text{since every subset } P \text{ of } \{1,2,\ldots,n\} \text{ satisfies} \\ \text{either } |P|=|Q| \text{ or } |P|\neq|Q| \text{ (but not both)}\end{array}\right)$$

$$= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)} + \underbrace{\sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|\neq|Q|}} 0}_{=0}$$

$$= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} (-1)^{\sigma}\prod_{i=1}^{n} a_{i,\sigma(i)}}_{\substack{=(-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A\right) \\ \text{(by Lemma 7.215)}}}$$

$$= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A\right).$$

This proves Theorem 6.156 **(b)**. □

Solution to Exercise 6.44. We have now proven both Lemma 6.158 and Theorem 6.156. Thus, Exercise 6.44 is solved. □

7.113. Solution to Exercise 6.45

Throughout this section, we shall use the notations introduced in Definition 6.78 and in Definition 6.153.

Before we start solving Exercise 6.45, let us show a trivial lemma:

Lemma 7.217. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let $\mathbf{i}$ be a finite list of elements of $\{1, 2, \ldots, n\}$. Let $\mathbf{j}$ be a finite list of elements of $\{1, 2, \ldots, m\}$. Then, $\left(\operatorname{sub}_{\mathbf{i}}^{\mathbf{j}} A\right)^T = \operatorname{sub}_{\mathbf{j}}^{\mathbf{i}} \left(A^T\right)$.

Proof of Lemma 7.217. Write the list $\mathbf{i}$ as $(i_1, i_2, \ldots, i_u)$. Write the list $\mathbf{j}$ as $(j_1, j_2, \ldots, j_v)$. Then, Lemma 7.217 follows immediately from Proposition 6.79 **(e)**. □

Corollary 7.218. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$.

Let A be an $n \times m$-matrix. Let U be a subset of $\{1, 2, \ldots, m\}$. Let V be a subset of $\{1, 2, \ldots, n\}$. Assume that $|U| = |V|$. Then,

$$\det\left(\operatorname{sub}_{w(U)}^{w(V)} \left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(V)}^{w(U)} A\right).$$

Proof of Corollary 7.218. Let $k = |U| = |V|$. Then, each of the lists $w(U)$ and $w(V)$ is a list of k elements. Hence, $\operatorname{sub}_{w(V)}^{w(U)} A$ is a $k \times k$-matrix. Thus, Exercise 6.4 (applied to k and $\operatorname{sub}_{w(V)}^{w(U)} A$ instead of n and A) yields

$$\det\left(\left(\operatorname{sub}_{w(V)}^{w(U)} A\right)^T\right) = \det\left(\operatorname{sub}_{w(V)}^{w(U)} A\right). \tag{1343}$$

But $w(U)$ is a list of elements of $\{1, 2, \ldots, m\}$ (since U is a subset of $\{1, 2, \ldots, m\}$). Similarly, $w(V)$ is a list of elements of $\{1, 2, \ldots, n\}$. Hence, Proposition 7.217 (applied to $\mathbf{i} = w(V)$ and $\mathbf{j} = w(U)$) yields $\left(\operatorname{sub}_{w(V)}^{w(U)} A\right)^T = \operatorname{sub}_{w(U)}^{w(V)} \left(A^T\right)$. Taking determinants on both sides of this equality, we obtain

$$\det\left(\left(\operatorname{sub}_{w(V)}^{w(U)} A\right)^T\right) = \det\left(\operatorname{sub}_{w(U)}^{w(V)} \left(A^T\right)\right).$$

Comparing this with (1343), we obtain $\det\left(\operatorname{sub}_{w(U)}^{w(V)} \left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(V)}^{w(U)} A\right)$. This proves Corollary 7.218. □

Solution to Exercise 6.45. **(a)** Let P be a subset of $\{1,2,\ldots,n\}$ satisfying $|P| = |R|$ and $P \neq R$.

Let $k = |P|$. The definition of $\widetilde{P}$ yields $\widetilde{P} = \{1,2,\ldots,n\} \setminus P$. Since P is a subset of $\{1,2,\ldots,n\}$, this leads to $\left|\widetilde{P}\right| = \underbrace{|\{1,2,\ldots,n\}|}_{=n} - \underbrace{|P|}_{=k} = n-k$. Thus, $n-k = \left|\widetilde{P}\right| \geq 0$, so that $n \geq k$ and thus $k \in \{0,1,\ldots,n\}$.

Let $[n]$ denote the set $\{1,2,\ldots,n\}$. Thus, $[n] = \{1,2,\ldots,n\}$.

We have $k = |P| = |R|$, so that $|R| = k$.

From $P \neq R$, we can easily deduce that $\widetilde{P} \cap R \neq \varnothing$ [592].

Let $(r_1, r_2, \ldots, r_k)$ be the list of all elements of R in increasing order (with no repetitions). (This is well-defined, because $|R| = k$.)

Let $(t_1, t_2, \ldots, t_{n-k})$ be the list of all elements of $\widetilde{P}$ in increasing order (with no repetitions). (This is well-defined, because $\left|\widetilde{P}\right| = n-k$.)

We know that $(r_1, r_2, \ldots, r_k)$ is a list of elements of R. Hence, the elements $r_1, r_2, \ldots, r_k$ belong to R, and thus to $[n]$ (since $R \subseteq \{1,2,\ldots,n\} = [n]$).

We know that $(t_1, t_2, \ldots, t_{n-k})$ is a list of elements of $\widetilde{P}$. Hence, the elements $t_1, t_2, \ldots, t_{n-k}$ belong to $\widetilde{P}$, and thus to $[n]$ (since $\widetilde{P} \subseteq \{1,2,\ldots,n\} = [n]$).

The n elements $r_1, r_2, \ldots, r_k, t_1, t_2, \ldots, t_{n-k}$ belong to $[n]$ (since the k elements $r_1, r_2, \ldots, r_k$ belong to $[n]$, and since the $n-k$ elements $t_1, t_2, \ldots, t_{n-k}$ belong to $[n]$). Hence, we can define an n-tuple $(\kappa_1, \kappa_2, \ldots, \kappa_n) \in [n]^n$ by

$$(\kappa_1, \kappa_2, \ldots, \kappa_n) = (r_1, r_2, \ldots, r_k, t_1, t_2, \ldots, t_{n-k}).$$

Consider this n-tuple. Due to its definition, we have

$$(\kappa_i = r_i \qquad \text{for every } i \in \{1,2,\ldots,k\}) \tag{1344}$$

and

$$(\kappa_i = t_{i-k} \qquad \text{for every } i \in \{k+1, k+2, \ldots, n\}). \tag{1345}$$

Define a map $\kappa : [n] \to [n]$ by

$$(\kappa(i) = \kappa_i \text{ for every } i \in [n]).$$

---

[592] *Proof.* Assume the contrary (for the sake of contradiction). Thus, $\widetilde{P} \cap R = \varnothing$.

Any three sets X, Y and Z satisfy $(X \cap Y) \setminus Z = X \cap (Y \setminus Z)$. Applying this to $X = R$, $Y = \{1,2,\ldots,n\}$ and $Z = P$, we obtain

$$(R \cap \{1,2,\ldots,n\}) \setminus P = R \cap \underbrace{(\{1,2,\ldots,n\} \setminus P)}_{=\widetilde{P}} = R \cap \widetilde{P} = \widetilde{P} \cap R = \varnothing.$$

Comparing this with $\underbrace{(R \cap \{1,2,\ldots,n\})}_{\substack{=R \\ (\text{since } R \subseteq \{1,2,\ldots,n\})}} \setminus P = R \setminus P$, we obtain $R \setminus P = \varnothing$. In other words, $R \subseteq P$. Combining this with $R \neq P$ (since $P \neq R$), we conclude that R is a proper subset of P.

But P is a finite set. Hence, every proper subset of P has a size strictly smaller than $|P|$. In other words, if Y is a proper subset of P, then $|Y| < |P|$. Applying this to $Y = R$, we conclude that $|R| < |P|$. Hence, $|R| < |P| = |R|$. This is absurd. This contradiction proves that our assumption was wrong, qed.

(This makes sense, since $\kappa_i \in [n]$ for every $i \in [n]$.) Then, every $i \in \{1, 2, \ldots, k\}$ satisfies

$$\kappa(i) = \kappa_i \qquad \text{(by the definition of } \kappa\text{)}$$
$$= r_i \qquad \text{(by (1344))}. \tag{1346}$$

Also, every $j \in \{1, 2, \ldots, n-k\}$ satisfies

$$\kappa(k+j) = \kappa_{k+j} \qquad \text{(by the definition of } \kappa\text{)}$$
$$= t_{(k+j)-k} \qquad \text{(by (1345), applied to } i = k+j\text{)}$$
$$= t_j. \tag{1347}$$

We have

$$w(R) = (\kappa(1), \kappa(2), \ldots, \kappa(k)) \tag{1348}$$

[593] and

$$w\left(\widetilde{P}\right) = (\kappa(k+1), \kappa(k+2), \ldots, \kappa(n)) \tag{1349}$$

[594].

Using $\widetilde{P} \cap R \neq \varnothing$, we can easily see that $\kappa \notin S_n$ [595].

[593] *Proof of (1348):* The lists $w(R)$ and $(r_1, r_2, \ldots, r_k)$ must be identical (since each of them is the list of all elements of R in increasing order (with no repetitions)). In other words, $w(R) = (r_1, r_2, \ldots, r_k)$. Now, (1346) shows that $\kappa(i) = r_i$ for every $i \in \{1, 2, \ldots, k\}$. In other words, $(\kappa(1), \kappa(2), \ldots, \kappa(k)) = (r_1, r_2, \ldots, r_k)$. Comparing this with $w(R) = (r_1, r_2, \ldots, r_k)$, we obtain $w(R) = (\kappa(1), \kappa(2), \ldots, \kappa(k))$. This proves (1348).

[594] *Proof of (1349):* The lists $w\left(\widetilde{P}\right)$ and $(t_1, t_2, \ldots, t_{n-k})$ must be identical (since each of them is the list of all elements of $\widetilde{P}$ in increasing order (with no repetitions)). In other words, $w\left(\widetilde{P}\right) = (t_1, t_2, \ldots, t_{n-k})$. Now, (1347) shows that $\kappa(k+i) = t_i$ for every $i \in \{1, 2, \ldots, n-k\}$. In other words, $(\kappa(k+1), \kappa(k+2), \ldots, \kappa(n)) = (t_1, t_2, \ldots, t_{n-k})$. Comparing this with $w\left(\widetilde{P}\right) = (t_1, t_2, \ldots, t_{n-k})$, we obtain $w\left(\widetilde{P}\right) = (\kappa(k+1), \kappa(k+2), \ldots, \kappa(n))$. This proves (1349).

[595] *Proof.* Assume the contrary (for the sake of contradiction). Thus, $\kappa \in S_n$. Hence, the inverse permutation $\kappa^{-1} \in S_n$ is well-defined.

Recall that $\widetilde{P} \cap R \neq \varnothing$. In other words, there exists a $\rho \in \widetilde{P} \cap R$. Consider this ρ.

Recall that $(r_1, r_2, \ldots, r_k)$ is a list of all elements of R. Hence, $\{r_1, r_2, \ldots, r_k\} = R$. Now, $\rho \in \widetilde{P} \cap R \subseteq R = \{r_1, r_2, \ldots, r_k\}$. In other words, there exists some $i \in \{1, 2, \ldots, k\}$ such that $\rho = r_i$. Consider this i.

Recall that $(t_1, t_2, \ldots, t_{n-k})$ is a list of all elements of $\widetilde{P}$. Hence, $\{t_1, t_2, \ldots, t_{n-k}\} = \widetilde{P}$. Now, $\rho \in \widetilde{P} \cap R \subseteq \widetilde{P} = \{t_1, t_2, \ldots, t_{n-k}\}$. In other words, there exists some $j \in \{1, 2, \ldots, n-k\}$ such that $\rho = t_j$. Consider this j.

The equality (1347) (applied to j instead of i) yields $\kappa(k+j) = t_j = \rho$ (since $\rho = t_j$). But from (1346), we obtain $\kappa(i) = r_i = \rho$ (since $\rho = r_i$). Comparing this with $\kappa(k+j) = \rho$, we obtain $\kappa(i) = \kappa(k+j)$. Now, $i = \kappa^{-1}\left(\underbrace{\kappa(i)}_{=\kappa(k+j)}\right) = \kappa^{-1}(\kappa(k+j)) = k + \underbrace{j}_{\substack{>0 \\ \text{(since } j \in \{1,2,\ldots,n-k\})}} > k$. But $i \in \{1, 2, \ldots, k\}$ and thus $i \leq k$. This contradicts $i > k$. This contradiction proves that our assumption was wrong. Qed.

Write the $n \times n$-matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Define an $n \times n$-matrix A_κ by $A_\kappa = \left(a_{\kappa(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Then, Lemma 6.17 **(b)** (applied to A, $a_{i,j}$ and A_κ instead of B, $b_{i,j}$ and B_κ) yields

$$\det\left(A_\kappa\right) = 0 \tag{1350}$$

(since $\kappa \notin S_n$).

Now, every subset Q of $\{1, 2, \ldots, n\}$ satisfies

$$\operatorname{sub}_{w(R)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} \left(A_\kappa\right) \tag{1351}$$

[596] and

$$\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} \left(A_\kappa\right) \tag{1353}$$

[597].

---

[596] *Proof of (1351):* Let Q be a subset of $\{1, 2, \ldots, n\}$. Write the list $w(Q)$ in the form $w(Q) = (q_1, q_2, \ldots, q_\ell)$ for some $\ell \in \mathbb{N}$.

From $w(R) = (\kappa(1), \kappa(2), \ldots, \kappa(k))$ and $w(Q) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{w(R)}^{w(Q)} A = \operatorname{sub}_{(\kappa(1),\kappa(2),\ldots,\kappa(k))}^{(q_1,q_2,\ldots,q_\ell)} A = \operatorname{sub}_{\kappa(1),\kappa(2),\ldots,\kappa(k)}^{q_1,q_2,\ldots,q_\ell} A = \left(a_{\kappa(x),q_y}\right)_{1\leq x\leq k,\ 1\leq y\leq \ell} \tag{1352}$$

$$\left(\text{by the definition of } \operatorname{sub}_{\kappa(1),\kappa(2),\ldots,\kappa(k)}^{q_1,q_2,\ldots,q_\ell} A, \text{ since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

On the other hand, from $w(Q) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} \left(A_\kappa\right) = \operatorname{sub}_{(1,2,\ldots,k)}^{(q_1,q_2,\ldots,q_\ell)} \left(A_\kappa\right) = \operatorname{sub}_{1,2,\ldots,k}^{q_1,q_2,\ldots,q_\ell} \left(A_\kappa\right) = \left(a_{\kappa(x),q_y}\right)_{1\leq x\leq k,\ 1\leq y\leq \ell}$$

(by the definition of $\operatorname{sub}_{1,2,\ldots,k}^{q_1,q_2,\ldots,q_\ell} \left(A_\kappa\right)$, since $A_\kappa = \left(a_{\kappa(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Comparing this with (1352), we obtain $\operatorname{sub}_{w(R)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)} \left(A_\kappa\right)$. This proves (1351).

[597] *Proof of (1353):* Let Q be a subset of $\{1, 2, \ldots, n\}$. Write the list $w\left(\widetilde{Q}\right)$ in the form $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$ for some $\ell \in \mathbb{N}$.

From $w\left(\widetilde{P}\right) = (\kappa(k+1), \kappa(k+2), \ldots, \kappa(n))$ and $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A = \operatorname{sub}_{(\kappa(k+1),\kappa(k+2),\ldots,\kappa(n))}^{(q_1,q_2,\ldots,q_\ell)} A = \operatorname{sub}_{\kappa(k+1),\kappa(k+2),\ldots,\kappa(n)}^{q_1,q_2,\ldots,q_\ell} A = \left(a_{\kappa(k+x),q_y}\right)_{1\leq x\leq n-k,\ 1\leq y\leq \ell} \tag{1354}$$

$$\left(\text{by the definition of } \operatorname{sub}_{\kappa(k+1),\kappa(k+2),\ldots,\kappa(n)}^{q_1,q_2,\ldots,q_\ell} A, \text{ since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).$$

On the other hand, from $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} \left(A_\kappa\right) = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{(q_1,q_2,\ldots,q_\ell)} \left(A_\kappa\right) = \operatorname{sub}_{k+1,k+2,\ldots,n}^{q_1,q_2,\ldots,q_\ell} \left(A_\kappa\right) = \left(a_{\kappa(k+x),q_y}\right)_{1\leq x\leq n-k,\ 1\leq y\leq \ell}$$

(by the definition of $\operatorname{sub}_{k+1,k+2,\ldots,n}^{q_1,q_2,\ldots,q_\ell} \left(A_\kappa\right)$, since $A_\kappa = \left(a_{\kappa(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Comparing this with (1354), we obtain $\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A = \operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})} \left(A_\kappa\right)$. This proves (1353).

Define a subset P' of $\{1,2,\ldots,n\}$ by $P' = \{1,2,\ldots,k\}$. Then, $|P'| = k = |P|$. We have $w(P') = (1,2,\ldots,k)$ [598] and $w\left(\widetilde{P'}\right) = (k+1,k+2,\ldots,n)$ [599].

---

[598] *Proof.* The definition of $w(P')$ yields

$$\begin{aligned} w(P') &= \left(\text{the list of all elements of } \underbrace{P'}_{=\{1,2,\ldots,k\}} \text{ in increasing order (with no repetitions)}\right) \\ &= (\text{the list of all elements of } \{1,2,\ldots,k\} \text{ in increasing order (with no repetitions)}) \\ &= (1,2,\ldots,k). \end{aligned}$$

Qed.

[599] *Proof.* We have $P' = \{1,2,\ldots,k\}$. Now,

$$\begin{aligned} \widetilde{P'} &= \{1,2,\ldots,n\} \setminus \underbrace{P'}_{=\{1,2,\ldots,k\}} \qquad \left(\text{by the definition of } \widetilde{P'}\right) \\ &= \{1,2,\ldots,n\} \setminus \{1,2,\ldots,k\} = \{k+1,k+2,\ldots,n\} \qquad (\text{since } k \in \{0,1,\ldots,n\}). \end{aligned}$$

Now, the definition of $w\left(\widetilde{P'}\right)$ yields

$$\begin{aligned} & w\left(\widetilde{P'}\right) \\ &= \left(\text{the list of all elements of } \underbrace{\widetilde{P'}}_{=\{k+1,k+2,\ldots,n\}} \text{ in increasing order (with no repetitions)}\right) \\ &= (\text{the list of all elements of } \{k+1,k+2,\ldots,n\} \text{ in increasing order (with no repetitions)}) \\ &= (k+1,k+2,\ldots,n). \end{aligned}$$

Qed.

Now, Theorem 6.156 **(a)** (applied to A_κ and P' instead of A and P) yields

$$\det(A_\kappa) = \underbrace{\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P'|}}}_{\substack{=\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \\ \text{(since } |P'|=|P|)}} (-1)^{\sum P'+\sum Q} \det\left(\underbrace{\operatorname{sub}_{w(P')}^{w(Q)}(A_\kappa)}_{\substack{=\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)}(A_\kappa) \\ \text{(since } w(P')=(1,2,\ldots,k))}}\right) \det\left(\underbrace{\operatorname{sub}_{w(\widetilde{P'})}^{w(\widetilde{Q})}(A_\kappa)}_{\substack{=\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})}(A_\kappa) \\ \text{(since } w(\widetilde{P'})=(k+1,k+2,\ldots,n))}}\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P'+\sum Q} \det\left(\underbrace{\operatorname{sub}_{(1,2,\ldots,k)}^{w(Q)}(A_\kappa)}_{\substack{=\operatorname{sub}_{w(R)}^{w(Q)}A \\ \text{(by (1351))}}}\right) \det\left(\underbrace{\operatorname{sub}_{(k+1,k+2,\ldots,n)}^{w(\widetilde{Q})}(A_\kappa)}_{\substack{=\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}A \\ \text{(by (1353))}}}\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P'+\sum Q} \det\left(\operatorname{sub}_{w(R)}^{w(Q)}A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}A\right).$$

Comparing this with (1350), we obtain

$$0 = \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P'+\sum Q} \det\left(\operatorname{sub}_{w(R)}^{w(Q)}A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}A\right).$$

Multiplying both sides of this equality by $(-1)^{\sum P-\sum P'}$, we obtain

$$0 = (-1)^{\sum P-\sum P'} \cdot \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P'+\sum Q} \det\left(\operatorname{sub}_{w(R)}^{w(Q)}A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}A\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \underbrace{(-1)^{\sum P-\sum P'}(-1)^{\sum P'+\sum Q}}_{\substack{=(-1)^{(\sum P-\sum P')+(\sum P'+\sum Q)} \\ =(-1)^{\sum P+\sum Q} \\ \text{(since } (\sum P-\sum P')+(\sum P'+\sum Q)=\sum P+\sum Q)}} \det\left(\operatorname{sub}_{w(R)}^{w(Q)}A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}A\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(R)}^{w(Q)}A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}A\right).$$

This solves Exercise 6.45 **(a)**.

(b) Let P be a subset of $\{1,2,\ldots,n\}$ satisfying $|P| = |R|$ and $P \neq R$. Every subset Q of $\{1,2,\ldots,n\}$ satisfying $|Q| = |P|$ satisfies

$$\det\left(\operatorname{sub}_{w(R)}^{w(Q)}\left(A^T\right)\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(Q)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right) \qquad (1355)$$

[600].

---

[600] *Proof of (1355):* Let Q be a subset of $\{1,2,\ldots,n\}$ satisfying $|Q| = |P|$. Then, $|Q| = |P| = |R|$. Thus, $|R| = |Q|$. Hence, Corollary 7.218 (applied to $m = n$, $U = R$ and $V = Q$) yields

$$\det\left(\operatorname{sub}_{w(R)}^{w(Q)}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(Q)}^{w(R)} A\right). \qquad (1356)$$

On the other hand, the definition of $\widetilde{P}$ yields $\widetilde{P} = \{1,2,\ldots,n\} \setminus P \subseteq \{1,2,\ldots,n\}$. Hence, $\widetilde{P}$ is a subset of $\{1,2,\ldots,n\}$. The same argument (applied to Q instead of P) shows that $\widetilde{Q}$ is a subset of $\{1,2,\ldots,n\}$. Furthermore, $\widetilde{P} = \{1,2,\ldots,n\} \setminus P$ and thus

$$\left|\widetilde{P}\right| = |\{1,2,\ldots,n\} \setminus P| = \underbrace{|\{1,2,\ldots,n\}|}_{=n} - |P| \qquad (\text{since } P \text{ is a subset of } \{1,2,\ldots,n\})$$
$$= n - |P|.$$

The same argument (applied to Q instead of P) shows that $\left|\widetilde{Q}\right| = n - |Q|$. Hence, $\left|\widetilde{Q}\right| = n - \underbrace{|Q|}_{=|P|} = n - |P|$. Comparing this with $\left|\widetilde{P}\right| = n - |P|$, we obtain $\left|\widetilde{P}\right| = \left|\widetilde{Q}\right|$. Thus, Corollary 7.218 (applied to $m = n$, $U = \widetilde{P}$ and $V = \widetilde{Q}$) yields

$$\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right). \qquad (1357)$$

Multiplying the equality (1356) with the equality (1357), we obtain

$$\det\left(\operatorname{sub}_{w(R)}^{w(Q)}\left(A^T\right)\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(Q)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right).$$

Thus, (1355) is proven.

Exercise 6.45 **(a)** (applied to A^T instead of A) yields

$$0 = \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \underbrace{(-1)^{\sum P+\sum Q}}_{\substack{=(-1)^{\sum Q+\sum P} \\ (\text{since } \sum P+\sum Q=\sum Q+\sum P)}} \underbrace{\det\left(\operatorname{sub}_{w(R)}^{w(Q)}\left(A^T\right)\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}\left(A^T\right)\right)}_{\substack{=\det\left(\operatorname{sub}_{w(Q)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right) \\ (\text{by (1355)})}}$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum Q+\sum P}\det\left(\operatorname{sub}_{w(Q)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right)$$

$$= \sum_{\substack{G\subseteq\{1,2,\ldots,n\};\\ |G|=|P|}} (-1)^{\sum G+\sum P}\det\left(\operatorname{sub}_{w(G)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{G})}^{w(\widetilde{P})} A\right) \qquad (1358)$$

(here, we have renamed the summation index Q as G).

Now, forget that we fixed P. We thus have proven that (1358) holds for every subset P of $\{1,2,\ldots,n\}$ satisfying $|P|=|R|$ and $P\neq R$.

Now, let Q be a subset of $\{1,2,\ldots,n\}$ satisfying $|Q|=|R|$ and $Q\neq R$. Then, we can apply (1358) to $P=Q$. We thus obtain

$$0 = \sum_{\substack{G\subseteq\{1,2,\ldots,n\};\\ |G|=|Q|}} (-1)^{\sum G+\sum Q}\det\left(\operatorname{sub}_{w(G)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{G})}^{w(\widetilde{Q})} A\right)$$

$$= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} A\right)$$

(here, we have renamed the summation index G as P). This solves Exercise 6.45 **(b)**. □

7.114. Solution to Exercise 6.46

Throughout this section, we shall use the notations introduced in Definition 6.78 and in Definition 6.153. Also, whenever m is an integer, we shall use the notation $[m]$ for the set $\{1,2,\ldots,m\}$.

Solution to Exercise 6.46. **(a)** Let $A\in\mathbb{K}^{m\times n}$.

Define a $p\in\mathbb{N}$ by $p=|J|$. From $|J|+|K|=n$, we obtain $|K|=n-\underbrace{|J|}_{=p}=n-p$. Hence, $n-p=|K|\geq 0$, so that $n\geq p$ and therefore $p\leq n$. Thus, $p\in\{0,1,\ldots,n\}$ and $\{1,2,\ldots,p\}\subseteq\{1,2,\ldots,n\}$ and $\{p+1,p+2,\ldots,n\}\subseteq\{1,2,\ldots,n\}$.

Let $(j_1,j_2,\ldots,j_p)$ be the list of all elements of J in increasing order (with no repetitions). (This is well-defined (by Definition 2.50), because $|J|=p$.)

Let $(k_1,k_2,\ldots,k_{n-p})$ be the list of all elements of K in increasing order (with no repetitions). (This is well-defined, because $|K|=n-p$.)

The p elements $j_1, j_2, \ldots, j_p$ all belong to J (since $\left(j_1, j_2, \ldots, j_p\right)$ is a list of elements of J), and thus to $[m]$ (since J is a subset of $\{1, 2, \ldots, m\} = [m]$). Similarly, the $n - p$ elements $k_1, k_2, \ldots, k_{n-p}$ belong to $[m]$. Combining the previous two sentences, we conclude that the $p + (n - p)$ elements $j_1, j_2, \ldots, j_p, k_1, k_2, \ldots, k_{n-p}$ belong to $[m]$. Hence, $\left(j_1, j_2, \ldots, j_p, k_1, k_2, \ldots, k_{n-p}\right) \in [m]^{p+(n-p)} = [m]^n$ (since $p + (n - p) = n$).

Thus, we can define an n-tuple $(\gamma_1, \gamma_2, \ldots, \gamma_n) \in [m]^n$ by

$$(\gamma_1, \gamma_2, \ldots, \gamma_n) = \left(j_1, j_2, \ldots, j_p, k_1, k_2, \ldots, k_{n-p}\right). \tag{1359}$$

Consider this $(\gamma_1, \gamma_2, \ldots, \gamma_n)$. From (1359), we obtain

$$\gamma_i = j_i \qquad \text{for each } i \in \{1, 2, \ldots, p\} \tag{1360}$$

and

$$\gamma_{p+i} = k_i \qquad \text{for each } i \in \{1, 2, \ldots, n - p\}. \tag{1361}$$

We have $\gamma_i \in [m]$ for each $i \in [n]$ (since $(\gamma_1, \gamma_2, \ldots, \gamma_n) \in [m]^n$). Hence, we can define a map $\gamma : [n] \to [m]$ by

$$(\gamma(i) = \gamma_i \text{ for every } i \in [n]).$$

Consider this γ.

We have

$$\gamma(i) = j_i \qquad \text{for every } i \in \{1, 2, \ldots, p\}. \tag{1362}$$

[*Proof of (1362):* Let $i \in \{1, 2, \ldots, p\}$. Then, $i \in \{1, 2, \ldots, p\} \subseteq \{1, 2, \ldots, n\} = [n]$. Hence, the definition of γ yields $\gamma(i) = \gamma_i = j_i$ (by (1360)). This proves (1362).]

Furthermore,

$$\gamma(p + i) = k_i \qquad \text{for every } i \in \{1, 2, \ldots, n - p\}. \tag{1363}$$

[*Proof of (1363):* Let $i \in \{1, 2, \ldots, n - p\}$. Thus, $p + i \in \{p + 1, p + 2, \ldots, n\} \subseteq \{1, 2, \ldots, n\} = [n]$. Hence, the definition of γ yields $\gamma(p + i) = \gamma_{p+i} = k_i$ (by (1361)). This proves (1363).]

We have

$$w(J) = (\gamma(1), \gamma(2), \ldots, \gamma(p)) \tag{1364}$$

[601] and

$$w(K) = (\gamma(p + 1), \gamma(p + 2), \ldots, \gamma(n)) \tag{1365}$$

[602].

[601] *Proof of (1364):* The lists $w(J)$ and $\left(j_1, j_2, \ldots, j_p\right)$ must be identical (since each of them is the list of all elements of J in increasing order (with no repetitions)). In other words, $w(J) = \left(j_1, j_2, \ldots, j_p\right)$. But (1362) yields $(\gamma(1), \gamma(2), \ldots, \gamma(p)) = \left(j_1, j_2, \ldots, j_p\right)$. Comparing this with $w(J) = \left(j_1, j_2, \ldots, j_p\right)$, we obtain $w(J) = (\gamma(1), \gamma(2), \ldots, \gamma(p))$.

[602] *Proof of (1365):* The lists $w(K)$ and $\left(k_1, k_2, \ldots, k_{n-p}\right)$ must be identical (since each of them is the list of all elements of K in increasing order (with no repetitions)). In other words, $w(K) = \left(k_1, k_2, \ldots, k_{n-p}\right)$. But (1362) yields $(\gamma(p + 1), \gamma(p + 2), \ldots, \gamma(n)) = \left(k_1, k_2, \ldots, k_{n-p}\right)$. Comparing this with $w(K) = \left(k_1, k_2, \ldots, k_{n-p}\right)$, we obtain $w(K) = (\gamma(p + 1), \gamma(p + 2), \ldots, \gamma(n))$.

Furthermore, there exist two distinct elements u and v of $[n]$ satisfying $\gamma(u) = \gamma(v)$ [603].

Write the $m \times n$-matrix A in the form $A = (a_{i,j})_{1\leq i\leq m,\ 1\leq j\leq n}$. Define an $n \times n$-matrix A_γ by $A_\gamma = \left(a_{\gamma(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Then, the matrix A_γ has two equal rows[604].

Hence, Exercise 6.7 **(e)** (applied to A_γ instead of A) shows that

$$\det(A_\gamma) = 0. \tag{1366}$$

Now, every subset Q of $\{1,2,\ldots,n\}$ satisfies

$$\operatorname{sub}_{w(J)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,p)}^{w(Q)} (A_\gamma) \tag{1367}$$

[605] and

$$\operatorname{sub}_{w(K)}^{w(\widetilde{Q})} A = \operatorname{sub}_{(p+1,p+2,\ldots,n)}^{w(\widetilde{Q})} (A_\gamma) \tag{1369}$$

---

[603] *Proof.* We have $J \cap K \neq \varnothing$. In other words, there exists some $h \in J \cap K$. Consider this h.

Recall that $(j_1, j_2, \ldots, j_p)$ is the list of all elements of J in increasing order (with no repetitions). Thus, $J = \{j_1, j_2, \ldots, j_p\}$. Similarly, $K = \{k_1, k_2, \ldots, k_{n-p}\}$.

We have $h \in J \cap K \subseteq J = \{j_1, j_2, \ldots, j_p\}$. In other words, there exists some $x \in \{1,2,\ldots,p\}$ such that $h = j_x$. Consider this x. We have $x \in \{1,2,\ldots,p\} \subseteq \{1,2,\ldots,n\} = [n]$.

We have $h \in J \cap K \subseteq K = \{k_1, k_2, \ldots, k_{n-p}\}$. In other words, there exists some $y \in \{1,2,\ldots,n-p\}$ such that $h = k_y$. Consider this y. We have $y \in \{1,2,\ldots,n-p\}$ and thus $p+y \in \{p+1, p+2, \ldots, n\} \subseteq \{1,2,\ldots,n\} = [n]$.

From $x \in \{1,2,\ldots,p\}$, we obtain $p \geq x$. From $y \in \{1,2,\ldots,n-p\}$, we obtain $y > 0$, so that $p + \underbrace{y}_{>0} > p \geq x$. Hence, $p + y \neq x$. Thus, the numbers x and $p+y$ are distinct.

Both x and $p+y$ are elements of $[n]$ (since $x \in [n]$ and $p+y \in [n]$). Hence, x and $p+y$ are two distinct elements of $[n]$.

Applying (1362) to $i = x$, we obtain $\gamma(x) = j_x = h$ (since $h = j_x$).

Applying (1363) to $i = y$, we obtain $\gamma(p+y) = k_y = h$ (since $h = k_y$).

Thus, $\gamma(x) = h = \gamma(p+y)$. Hence, there exist two distinct elements u and v of $[n]$ satisfying $\gamma(u) = \gamma(v)$ (namely, $u = x$ and $v = p+y$).

[604] *Proof.* We have shown that there exist two distinct elements u and v of $[n]$ satisfying $\gamma(u) = \gamma(v)$. Consider these u and v.

We have $A_\gamma = \left(a_{\gamma(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Thus,

$$(\text{the } u\text{-th row of the matrix } A_\gamma) = \left(a_{\gamma(u),1}, a_{\gamma(u),2}, \ldots, a_{\gamma(u),n}\right) = \left(a_{\gamma(v),1}, a_{\gamma(v),2}, \ldots, a_{\gamma(v),n}\right)$$

(since $\gamma(u) = \gamma(v)$). Comparing this with

$$\begin{aligned}&(\text{the } v\text{-th row of the matrix } A_\gamma)\\ &= \left(a_{\gamma(v),1}, a_{\gamma(v),2}, \ldots, a_{\gamma(v),n}\right) \qquad \left(\text{since } A_\gamma = \left(a_{\gamma(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right),\end{aligned}$$

we obtain (the u-th row of the matrix A_γ) = (the v-th row of the matrix A_γ). Since u and v are distinct, this shows that the matrix A_γ has two equal rows.

[605] *Proof of (1367):* Let Q be a subset of $\{1,2,\ldots,n\}$. Write the list $w(Q)$ in the form $w(Q) = (q_1, q_2, \ldots, q_\ell)$ for some $\ell \in \mathbb{N}$.

[606].

But $\{1,2,\ldots,p\} \subseteq \{1,2,\ldots,n\}$. In other words, $\{1,2,\ldots,p\}$ is a subset of $\{1,2,\ldots,n\}$. Thus, we can define a subset P' of $\{1,2,\ldots,n\}$ by $P' = \{1,2,\ldots,p\}$. Consider this P'. From $P' = \{1,2,\ldots,p\}$, we obtain $|P'| = |\{1,2,\ldots,p\}| = p = |J|$.

From $w(J) = (\gamma(1), \gamma(2), \ldots, \gamma(p))$ and $w(Q) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{w(J)}^{w(Q)} A = \operatorname{sub}_{(\gamma(1),\gamma(2),\ldots,\gamma(p))}^{(q_1,q_2,\ldots,q_\ell)} A = \operatorname{sub}_{\gamma(1),\gamma(2),\ldots,\gamma(p)}^{q_1,q_2,\ldots,q_\ell} A = \left(a_{\gamma(x),q_y}\right)_{1\leq x\leq p,\ 1\leq y\leq \ell} \tag{1368}$$

$$\left(\text{by the definition of } \operatorname{sub}_{\gamma(1),\gamma(2),\ldots,\gamma(p)}^{q_1,q_2,\ldots,q_\ell} A, \text{ since } A = \left(a_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq n}\right).$$

On the other hand, from $w(Q) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{(1,2,\ldots,p)}^{w(Q)} (A_\gamma) = \operatorname{sub}_{(1,2,\ldots,p)}^{(q_1,q_2,\ldots,q_\ell)} (A_\gamma) = \operatorname{sub}_{1,2,\ldots,p}^{q_1,q_2,\ldots,q_\ell} (A_\gamma) = \left(a_{\gamma(x),q_y}\right)_{1\leq x\leq p,\ 1\leq y\leq \ell}$$

(by the definition of $\operatorname{sub}_{1,2,\ldots,p}^{q_1,q_2,\ldots,q_\ell} (A_\gamma)$, since $A_\gamma = \left(a_{\gamma(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Comparing this with (1368), we obtain $\operatorname{sub}_{w(J)}^{w(Q)} A = \operatorname{sub}_{(1,2,\ldots,p)}^{w(Q)} (A_\gamma)$. This proves (1367).

[606] *Proof of (1369):* Let Q be a subset of $\{1,2,\ldots,n\}$. Write the list $w\left(\widetilde{Q}\right)$ in the form $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$ for some $\ell \in \mathbb{N}$.

From $w(K) = (\gamma(p+1), \gamma(p+2), \ldots, \gamma(n))$ and $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{w(K)}^{w(\widetilde{Q})} A = \operatorname{sub}_{(\gamma(p+1),\gamma(p+2),\ldots,\gamma(n))}^{(q_1,q_2,\ldots,q_\ell)} A = \operatorname{sub}_{\gamma(p+1),\gamma(p+2),\ldots,\gamma(n)}^{q_1,q_2,\ldots,q_\ell} A = \left(a_{\gamma(p+x),q_y}\right)_{1\leq x\leq n-p,\ 1\leq y\leq \ell} \tag{1370}$$

$$\left(\text{by the definition of } \operatorname{sub}_{\gamma(p+1),\gamma(p+2),\ldots,\gamma(n)}^{q_1,q_2,\ldots,q_\ell} A, \text{ since } A = \left(a_{i,j}\right)_{1\leq i\leq m,\ 1\leq j\leq n}\right).$$

On the other hand, from $w\left(\widetilde{Q}\right) = (q_1, q_2, \ldots, q_\ell)$, we obtain

$$\operatorname{sub}_{(p+1,p+2,\ldots,n)}^{w(\widetilde{Q})} (A_\gamma) = \operatorname{sub}_{(p+1,p+2,\ldots,n)}^{(q_1,q_2,\ldots,q_\ell)} (A_\gamma) = \operatorname{sub}_{p+1,p+2,\ldots,n}^{q_1,q_2,\ldots,q_\ell} (A_\gamma) = \left(a_{\gamma(p+x),q_y}\right)_{1\leq x\leq n-p,\ 1\leq y\leq \ell}$$

(by the definition of $\operatorname{sub}_{p+1,p+2,\ldots,n}^{q_1,q_2,\ldots,q_\ell} (A_\gamma)$, since $A_\gamma = \left(a_{\gamma(i),j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Comparing this with (1370), we obtain $\operatorname{sub}_{w(K)}^{w(\widetilde{Q})} A = \operatorname{sub}_{(p+1,p+2,\ldots,n)}^{w(\widetilde{Q})} (A_\gamma)$. This proves (1369).

We have $w(P') = (1, 2, \ldots, p)$ [607] and $w\left(\widetilde{P'}\right) = (p+1, p+2, \ldots, n)$ [608].

[607] *Proof.* Recall that $w(P')$ is the list of all elements of P' in increasing order (with no repetitions) (by the definition of $w(P')$). Thus,

$$\begin{aligned}
w\left(P'\right) &= \left(\text{the list of all elements of } \underbrace{P'}_{=\{1,2,\ldots,p\}} \text{ in increasing order (with no repetitions)}\right) \\
&= \left(\text{the list of all elements of } \{1,2,\ldots,p\} \text{ in increasing order (with no repetitions)}\right) \\
&= (1,2,\ldots,p).
\end{aligned}$$

Qed.

[608] *Proof.* We have

$$\begin{aligned}
\widetilde{P'} &= \{1,2,\ldots,n\} \setminus \underbrace{P'}_{=\{1,2,\ldots,p\}} \qquad \left(\text{by the definition of } \widetilde{P'}\right) \\
&= \{1,2,\ldots,n\} \setminus \{1,2,\ldots,p\} = \{p+1, p+2, \ldots, n\} \qquad (\text{since } p \in \{0,1,\ldots,n\}).
\end{aligned}$$

Recall that $w\left(\widetilde{P'}\right)$ is the list of all elements of $\widetilde{P'}$ in increasing order (with no repetitions) (by the definition of $w\left(\widetilde{P'}\right)$). Thus,

$$\begin{aligned}
& w\left(\widetilde{P'}\right) \\
&= \left(\text{the list of all elements of } \underbrace{\widetilde{P'}}_{=\{p+1,p+2,\ldots,n\}} \text{ in increasing order (with no repetitions)}\right) \\
&= \left(\text{the list of all elements of } \{p+1,p+2,\ldots,n\} \text{ in increasing order (with no repetitions)}\right) \\
&= (p+1, p+2, \ldots, n).
\end{aligned}$$

Qed.

Now, Theorem 6.156 **(a)** (applied to A_γ and P' instead of A and P) yields

$$\det\left(A_\gamma\right) = \underbrace{\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P'|}}}_{\substack{=\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} \\ (\text{since } |P'|=|J|)}} (-1)^{\sum P'+\sum Q} \det\left(\underbrace{\operatorname{sub}_{w(P')}^{w(Q)}\left(A_\gamma\right)}_{\substack{=\operatorname{sub}_{(1,2,\ldots,p)}^{w(Q)}(A_\gamma) \\ (\text{since } w(P')=(1,2,\ldots,p))}}\right) \det\left(\underbrace{\operatorname{sub}_{w\left(\widetilde{P'}\right)}^{w\left(\widetilde{Q}\right)}\left(A_\gamma\right)}_{\substack{=\operatorname{sub}_{(p+1,p+2,\ldots,n)}^{w(\widetilde{Q})}(A_\gamma) \\ (\text{since } w\left(\widetilde{P'}\right)=(p+1,p+2,\ldots,n))}}\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} \underbrace{(-1)^{\sum P'+\sum Q}}_{=(-1)^{\sum P'}(-1)^{\sum Q}} \det\left(\underbrace{\operatorname{sub}_{(1,2,\ldots,p)}^{w(Q)}\left(A_\gamma\right)}_{\substack{=\operatorname{sub}_{w(J)}^{w(Q)} A \\ (\text{by } (1367))}}\right) \det\left(\underbrace{\operatorname{sub}_{(p+1,p+2,\ldots,n)}^{w\left(\widetilde{Q}\right)}\left(A_\gamma\right)}_{\substack{=\operatorname{sub}_{w(K)}^{w(\widetilde{Q})} A \\ (\text{by } (1369))}}\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} (-1)^{\sum P'} (-1)^{\sum Q} \det\left(\operatorname{sub}_{w(J)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(K)}^{w\left(\widetilde{Q}\right)} A\right).$$

Comparing this with (1366), we obtain

$$0 = \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} (-1)^{\sum P'} (-1)^{\sum Q} \det\left(\operatorname{sub}_{w(J)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(K)}^{w\left(\widetilde{Q}\right)} A\right).$$

Multiplying both sides of this equality by $(-1)^{\sum P'}$, we obtain

$$0 = (-1)^{\sum P'} \cdot \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} (-1)^{\sum P'} (-1)^{\sum Q} \det\left(\operatorname{sub}_{w(J)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(K)}^{w\left(\widetilde{Q}\right)} A\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} \underbrace{(-1)^{\sum P'} (-1)^{\sum P'}}_{\substack{=(-1)^{\sum P'+\sum P'}=1 \\ (\text{since } \sum P'+\sum P'=2\sum P' \text{ is even})}} (-1)^{\sum Q} \det\left(\operatorname{sub}_{w(J)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(K)}^{w\left(\widetilde{Q}\right)} A\right)$$

$$= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|J|}} (-1)^{\sum Q} \det\left(\operatorname{sub}_{w(J)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(K)}^{w\left(\widetilde{Q}\right)} A\right).$$

This solves Exercise 6.46 **(a)**.

(b) Let $A \in \mathbb{K}^{n \times m}$. Every subset Q of $\{1, 2, \ldots, n\}$ satisfying $|Q| = |J|$ satisfies

$$\det\left(\operatorname{sub}_{w(J)}^{w(Q)}\left(A^T\right)\right) \det\left(\operatorname{sub}_{w(K)}^{w(\widetilde{Q})}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(Q)}^{w(J)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(K)} A\right) \qquad (1371)$$

[609].

We have $A \in \mathbb{K}^{n \times m}$, and thus $A^T \in \mathbb{K}^{m \times n}$. Hence, Exercise 6.46 **(a)** (applied to A^T instead of A) yields

$$\begin{aligned} 0 &= \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |Q| = |J|}} (-1)^{\sum Q} \underbrace{\det\left(\operatorname{sub}_{w(J)}^{w(Q)}\left(A^T\right)\right) \det\left(\operatorname{sub}_{w(K)}^{w(\widetilde{Q})}\left(A^T\right)\right)}_{\substack{= \det\left(\operatorname{sub}_{w(Q)}^{w(J)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(K)} A\right) \\ \text{(by (1371))}}} \\ &= \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |Q| = |J|}} (-1)^{\sum Q} \det\left(\operatorname{sub}_{w(Q)}^{w(J)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(K)} A\right) \\ &= \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |J|}} (-1)^{\sum P} \det\left(\operatorname{sub}_{w(P)}^{w(J)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(K)} A\right) \end{aligned} \qquad (1374)$$

(here, we have renamed the summation index Q as P). This solves Exercise 6.46 **(b)**. $\square$

[609] *Proof of (1371):* Let Q be a subset of $\{1, 2, \ldots, n\}$ satisfying $|Q| = |J|$. Thus, $|J| = |Q|$. Hence, Corollary 7.218 (applied to $U = J$ and $V = Q$) yields

$$\det\left(\operatorname{sub}_{w(J)}^{w(Q)}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(Q)}^{w(J)} A\right). \qquad (1372)$$

On the other hand, the definition of $\widetilde{Q}$ yields $\widetilde{Q} = \{1, 2, \ldots, n\} \setminus Q \subseteq \{1, 2, \ldots, n\}$. Hence, $\widetilde{Q}$ is a subset of $\{1, 2, \ldots, n\}$. Also, $\widetilde{Q} = \{1, 2, \ldots, n\} \setminus Q$ and thus

$$\begin{aligned} \left|\widetilde{Q}\right| &= |\{1, 2, \ldots, n\} \setminus Q| = \underbrace{|\{1, 2, \ldots, n\}|}_{=n} - \underbrace{|Q|}_{=|J|} \qquad (\text{since } Q \text{ is a subset of } \{1, 2, \ldots, n\}) \\ &= n - |J| = |K| \qquad (\text{since } n = |J| + |K|). \end{aligned}$$

Hence, $|K| = \left|\widetilde{Q}\right|$. Thus, Corollary 7.218 (applied to $U = K$ and $V = \widetilde{Q}$) yields

$$\det\left(\operatorname{sub}_{w(K)}^{w(\widetilde{Q})}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(K)} A\right). \qquad (1373)$$

Multiplying the equality (1372) with the equality (1373), we obtain

$$\det\left(\operatorname{sub}_{w(J)}^{w(Q)}\left(A^T\right)\right) \det\left(\operatorname{sub}_{w(K)}^{w(\widetilde{Q})}\left(A^T\right)\right) = \det\left(\operatorname{sub}_{w(Q)}^{w(J)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(K)} A\right).$$

Thus, (1371) is proven.

7.115. Solution to Exercise 6.47

Before we solve Exercise 6.47 using Theorem 6.156, let us outline a quick solution for Exercise 6.47 **(a)** using just the definition of the determinant (generalizing our first solution to Exercise 6.6 **(a)**):

*First solution to Exercise 6.47 **(a)** (sketched).* Assume that $|P| + |Q| > n$.

Let $\sigma \in S_n$. We shall write $[n]$ for the set $\{1, 2, \ldots, n\}$.

Assume (for the sake of contradiction) that $\sigma(P) \subseteq [n] \setminus Q$. Thus,

$$|\sigma(P)| \leq |[n] \setminus Q| = \underbrace{|[n]|}_{=n} - |Q| \qquad (\text{since } Q \subseteq [n])$$
$$= n - |Q| < |P| \qquad (\text{since } |P| + |Q| > n).$$

But the map σ is injective (since it is a permutation); thus, $|\sigma(P)| = |P|$. Hence, $|P| = |\sigma(P)| < |P|$, which is absurd. This contradiction shows that our assumption (that $\sigma(P) \subseteq [n] \setminus Q$) was false. Hence, $\sigma(P) \not\subseteq [n] \setminus Q$. In other words, there exists some $p \in P$ such that $\sigma(p) \notin [n] \setminus Q$. Consider this p.

From $\sigma(p) \in [n]$ and $\sigma(p) \notin [n] \setminus Q$, we obtain $\sigma(p) \in [n] \setminus ([n] \setminus Q) \subseteq Q$. Hence, (473) (applied to $i = p$ and $j = \sigma(p)$) yields $a_{p,\sigma(p)} = 0$.

Now, $a_{p,\sigma(p)}$ is a factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$. Thus, at least one factor of the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ is 0 (namely, $a_{p,\sigma(p)} = 0$). Hence, the whole product must be 0. In other words, $\prod_{i=1}^{n} a_{i,\sigma(i)} = 0$.

Now, forget that we fixed σ. We thus have shown that $\prod_{i=1}^{n} a_{i,\sigma(i)} = 0$ for each $\sigma \in S_n$. Hence, (341) becomes

$$\det A = \sum_{\sigma \in S_n} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n} a_{i,\sigma(i)}}_{=0} = \sum_{\sigma \in S_n} (-1)^{\sigma} 0 = 0.$$

This solves Exercise 6.47 **(a)**. □

Let us now prepare for a "proper" solution to Exercise 6.47, which will solve both parts **(a)** and **(b)**. We shall use the notations introduced in Definition 6.78 and in Definition 6.153. We shall also use the following lemma:

Lemma 7.219. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. (For instance, if $n = 4$ and $I = \{1, 4\}$, then $\widetilde{I} = \{2, 3\}$.)

Let P and Q be two subsets of $\{1, 2, \ldots, n\}$. Let R be a subset of $\{1, 2, \ldots, n\}$ satisfying $|R| = |P|$ and $R \not\subseteq \widetilde{Q}$.

Let $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n} \in \mathbb{K}^{n\times n}$ be an $n \times n$-matrix such that

$$\text{every } i \in P \text{ and } j \in Q \text{ satisfy } a_{i,j} = 0. \tag{1375}$$

Then, $\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) = 0$.

Proof of Lemma 7.219. Define $k \in \mathbb{N}$ by $k = |R| = |P|$. (This is well-defined, since $|R| = |P|$.)

Recall that $w(P)$ is the list of all elements of P in increasing order (with no repetitions) (by the definition of $w(P)$). Thus, $w(P)$ is a list of k elements (since $|P| = k$). Write this list $w(P)$ in the form $w(P) = (p_1, p_2, \ldots, p_k)$. Thus, $P = \{p_1, p_2, \ldots, p_k\}$.

Similarly, write the list $w(R)$ in the form $w(R) = (r_1, r_2, \ldots, r_k)$. Thus, $R = \{r_1, r_2, \ldots, r_k\}$.

We have $R \not\subseteq \widetilde{Q}$. In other words, there exists some $z \in R$ such that $z \notin \widetilde{Q}$. Consider this z. We have $\widetilde{Q} = \{1, 2, \ldots, n\} \setminus Q$ (by the definition of $\widetilde{Q}$). Combining $z \in R \subseteq \{1, 2, \ldots, n\}$ with $z \notin \widetilde{Q}$, we obtain

$$z \in \{1, 2, \ldots, n\} \setminus \underbrace{\widetilde{Q}}_{=\{1,2,\ldots,n\}\setminus Q} = \{1, 2, \ldots, n\} \setminus (\{1, 2, \ldots, n\} \setminus Q) \subseteq Q.$$

But $z \in R = \{r_1, r_2, \ldots, r_k\}$. Hence, $z = r_v$ for some $v \in \{1, 2, \ldots, k\}$. Consider this v. We have

$$a_{p_x, r_v} = 0 \qquad \text{for each } x \in \{1, 2, \ldots, k\}. \tag{1376}$$

[*Proof of (1376):* Let $x \in \{1, 2, \ldots, k\}$. Then, $p_x \in \{p_1, p_2, \ldots, p_k\} = P$. Also, $r_v = z \in Q$. Hence, (1375) (applied to $i = p_x$ and $j = r_v$) yields $a_{p_x, r_v} = 0$. This proves (1376).]

From $w(P) = (p_1, p_2, \ldots, p_k)$ and $w(R) = (r_1, r_2, \ldots, r_k)$, we obtain

$$\operatorname{sub}_{w(P)}^{w(R)} A = \operatorname{sub}_{(p_1,p_2,\ldots,p_k)}^{(r_1,r_2,\ldots,r_k)} A = \operatorname{sub}_{p_1,p_2,\ldots,p_k}^{r_1,r_2,\ldots,r_k} A = \left(a_{p_x, r_y}\right)_{1\leq x\leq k,\ 1\leq y\leq k}.$$

Hence, $\operatorname{sub}_{w(P)}^{w(R)} A$ is a $k \times k$-matrix, and its v-th column is

$$\begin{pmatrix} a_{p_1,r_v} \\ a_{p_2,r_v} \\ \vdots \\ a_{p_k,r_v} \end{pmatrix} = \left(\underbrace{a_{p_x,r_v}}_{\substack{=0 \\ \text{(by (1376))}}} \right)_{1\leq x\leq k,\ 1\leq y\leq 1} = (0)_{1\leq x\leq k,\ 1\leq y\leq 1}.$$

In other words, the v-th column of the matrix $\operatorname{sub}_{w(P)}^{w(R)} A$ consists of zeroes. Thus, a column of the matrix $\operatorname{sub}_{w(P)}^{w(R)} A$ consists of zeroes.

Hence, Exercise 6.7 **(d)** (applied to k and $\operatorname{sub}_{w(P)}^{w(R)} A$ instead of n and A) yields $\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) = 0$. This proves Lemma 7.219. $\square$

Solution to Exercise 6.47. For any subset I of $\{1,2,\ldots,n\}$, we let $\widetilde{I}$ denote the complement $\{1,2,\ldots,n\} \setminus I$ of I. (For instance, if $n=4$ and $I=\{1,4\}$, then $\widetilde{I}=\{2,3\}$.)

Thus, $\widetilde{Q}$ is the complement of Q. Hence, Q is, in turn, the complement of $\widetilde{Q}$. In other words, $Q=\widetilde{\widetilde{Q}}$. Also, $\widetilde{Q}$ is the complement of Q in the n-element set $\{1,2,\ldots,n\}$; therefore, $\left|\widetilde{Q}\right| = n - |Q|$.

For every subset U of $\{1,2,\ldots,n\}$, we have

$$\det A = \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=|U|}} (-1)^{\sum U+\sum R} \det\left(\operatorname{sub}_{w(U)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{U})}^{w(\widetilde{R})} A\right). \qquad (1377)$$

(Indeed, this is merely the claim of Theorem 6.156 **(a)**, with the variables P and Q renamed as U and R.)

(a) Assume that $|P|+|Q|>n$. Thus, $|P|>n-|Q|=\left|\widetilde{Q}\right|$.

Hence, each subset R of $\{1,2,\ldots,n\}$ satisfying $|R|=|P|$ must satisfy

$$\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) = 0. \qquad (1378)$$

[*Proof of (1378):* Let R be a subset of $\{1,2,\ldots,n\}$ satisfying $|R|=|P|$. Then, $|R|=|P|>\left|\widetilde{Q}\right|$. If we had $R\subseteq\widetilde{Q}$, then we would have $|R|\leq\left|\widetilde{Q}\right|$, which would contradict $|R|>\left|\widetilde{Q}\right|$. Thus, we cannot have $R\subseteq\widetilde{Q}$. Hence, we have $R\not\subseteq\widetilde{Q}$. Hence, Lemma 7.219 yields $\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right)=0$. This proves (1378).]

Now, (1377) (applied to $U=P$) yields

$$\det A = \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=|P|}} (-1)^{\sum P+\sum R} \underbrace{\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right)}_{\substack{=0\\ \text{(by (1378))}}} \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{R})} A\right)$$
$$= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=|P|}} (-1)^{\sum P+\sum R} 0 \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{R})} A\right) = 0.$$

This solves Exercise 6.47 **(a)**.

(b) Assume that $|P|+|Q|=n$. Thus, $|P|=n-|Q|=\left|\widetilde{Q}\right|$. In other words, $\left|\widetilde{Q}\right|=|P|$. Hence, $\widetilde{Q}$ is a subset R of $\{1,2,\ldots,n\}$ satisfying $|R|=|P|$ (since $\widetilde{Q}$ is a subset of $\{1,2,\ldots,n\}$).

Hence, each subset R of $\{1,2,\ldots,n\}$ satisfying $|R|=|P|$ and $R\neq\widetilde{Q}$ must satisfy

$$\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) = 0. \qquad (1379)$$

[*Proof of (1379):* Let R be a subset of $\{1,2,\ldots,n\}$ satisfying $|R|=|P|$ and $R\neq\widetilde{Q}$. Thus, $|R|=|P|=\left|\widetilde{Q}\right|$.

Thus, the set R has the same size as $\widetilde{Q}$. If R was a subset of $\widetilde{Q}$, then this would lead to $R = \widetilde{Q}$ (because the only subset of $\widetilde{Q}$ having the same size as $\widetilde{Q}$ is $\widetilde{Q}$ itself), which would contradict $R \neq \widetilde{Q}$. Hence, R is not a subset of $\widetilde{Q}$.

In other words, $R \not\subseteq \widetilde{Q}$. Hence, Lemma 7.219 yields $\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) = 0$. This proves (1379).]

Now, (1377) (applied to $U = P$) yields

$$\begin{aligned}
\det A &= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=|P|}} (-1)^{\sum P+\sum R} \det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{R})} A\right) \\
&= (-1)^{\sum P+\sum \widetilde{Q}} \det\left(\operatorname{sub}_{w(P)}^{w(\widetilde{Q})} A\right) \underbrace{\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{\widetilde{Q}})} A\right)}_{\substack{=\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(Q)} A\right)\\ (\text{since } \widetilde{\widetilde{Q}}=Q)}} \\
&\qquad + \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=|P|;\\ R\neq\widetilde{Q}}} (-1)^{\sum P+\sum R} \underbrace{\det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right)}_{\substack{=0\\ (\text{by (1379)})}} \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{R})} A\right) \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } R=\widetilde{Q} \text{ from the sum}\\ (\text{since } \widetilde{Q} \text{ is a subset } R \text{ of } \{1,2,\ldots,n\} \text{ satisfying } |R|=|P|)\end{array}\right) \\
&= (-1)^{\sum P+\sum \widetilde{Q}} \det\left(\operatorname{sub}_{w(P)}^{w(\widetilde{Q})} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(Q)} A\right) \\
&\qquad + \underbrace{\sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=|P|;\\ R\neq\widetilde{Q}}} (-1)^{\sum P+\sum R} 0 \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{R})} A\right)}_{=0} \\
&= (-1)^{\sum P+\sum \widetilde{Q}} \det\left(\operatorname{sub}_{w(P)}^{w(\widetilde{Q})} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(Q)} A\right).
\end{aligned}$$

This solves Exercise 6.47 **(b)**. □

7.116. Solution to Exercise 6.48

Proof of Theorem 6.160. Write the $n \times n$-matrix A in the form $A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. Write the $n \times n$-matrix B in the form $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$.

Let $[n]$ denote the set $\{1,2,\ldots,n\}$. Then, every subset I of $[n]$ satisfies

$$[n] \setminus I = \widetilde{I} \tag{1380}$$

(because the definition of $\widetilde{I}$ yields $\widetilde{I} = \underbrace{\{1, 2, \ldots, n\}}_{=[n]} \setminus I = [n] \setminus I$).

For every $\sigma \in S_n$ and every subset P of $\{1, 2, \ldots, n\}$, define an element $c_{\sigma,P}$ of $\mathbb{K}$ by

$$c_{\sigma,P} = (-1)^{\sigma} \left(\prod_{i \in P} a_{i,\sigma(i)} \right) \left(\prod_{i \in \widetilde{P}} b_{i,\sigma(i)} \right). \tag{1381}$$

If P and Q are two subsets of $\{1, 2, \ldots, n\}$ satisfying $|Q| \neq |P|$, then

$$\sum_{\substack{\sigma \in S_n; \\ \sigma(P) = Q}} c_{\sigma,P} = 0 \tag{1382}$$

610.

If P and Q are two subsets of $\{1, 2, \ldots, n\}$ satisfying $|Q| = |P|$, then

$$\sum_{\substack{\sigma \in S_n; \\ \sigma(P) = Q}} c_{\sigma,P} = (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B \right) \tag{1383}$$

611.

Adding the equalities $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ and $B = \left(b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, we obtain $A + B = \left(a_{i,j} + b_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus, (341) (applied to $A + B$ and $a_{i,j} + b_{i,j}$ instead

[610] *Proof of (1382):* This can be shown in precisely the same way as (1342) was shown in our proof of Theorem 6.156.

[611] *Proof of (1383):* Let P and Q be two subsets of $\{1, 2, \ldots, n\}$ satisfying $|Q| = |P|$. From $|Q| = |P|$, we obtain $|P| = |Q|$. Lemma 6.158 thus yields

$$\sum_{\substack{\sigma \in S_n; \\ \sigma(P) = Q}} (-1)^{\sigma} \left(\prod_{i \in P} a_{i,\sigma(i)} \right) \left(\prod_{i \in \widetilde{P}} b_{i,\sigma(i)} \right) = (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B \right).$$

But this is precisely the equality (1383) (because of (1381)). Thus, (1383) is proven.

of A and $a_{i,j}$) yields

$$\det\left(A+B\right) = \sum_{\sigma\in S_n}\left(-1\right)^{\sigma} \underbrace{\prod_{i=1}^{n}\left(a_{i,\sigma(i)}+b_{i,\sigma(i)}\right)}_{\substack{=\sum_{I\subseteq[n]}\left(\prod_{i\in I}a_{i,\sigma(i)}\right)\left(\prod_{i\in[n]\setminus I}b_{i,\sigma(i)}\right) \\ \text{(by Exercise 6.1 \textbf{(a)},} \\ \text{applied to } a_{i,\sigma(i)} \text{ and } b_{i,\sigma(i)} \text{ instead of } a_i \text{ and } b_i\text{)}}}$$

$$= \sum_{\sigma\in S_n}\left(-1\right)^{\sigma} \underbrace{\sum_{I\subseteq[n]}}_{=\sum_{I\subseteq\{1,2,\ldots,n\}}} \left(\prod_{i\in I}a_{i,\sigma(i)}\right)\left(\underbrace{\prod_{i\in[n]\setminus I}}_{\substack{=\prod_{i\in\widetilde{I}} \\ \text{(by (1380))}}} b_{i,\sigma(i)}\right)$$

$$= \sum_{\sigma\in S_n}\left(-1\right)^{\sigma} \underbrace{\sum_{I\subseteq\{1,2,\ldots,n\}}\left(\prod_{i\in I}a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{I}}b_{i,\sigma(i)}\right)}_{\substack{=\sum_{P\subseteq\{1,2,\ldots,n\}}\left(\prod_{i\in P}a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{P}}b_{i,\sigma(i)}\right) \\ \text{(here, we have renamed the summation index } I \text{ as } P\text{)}}}$$

$$= \sum_{\sigma\in S_n}\left(-1\right)^{\sigma} \sum_{P\subseteq\{1,2,\ldots,n\}}\left(\prod_{i\in P}a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{P}}b_{i,\sigma(i)}\right)$$

$$= \sum_{P\subseteq\{1,2,\ldots,n\}}\sum_{\sigma\in S_n} \underbrace{\left(-1\right)^{\sigma}\left(\prod_{i\in P}a_{i,\sigma(i)}\right)\left(\prod_{i\in\widetilde{P}}b_{i,\sigma(i)}\right)}_{\substack{=c_{\sigma,P} \\ \text{(by (1381))}}}$$

$$= \sum_{P\subseteq\{1,2,\ldots,n\}}\sum_{\sigma\in S_n} c_{\sigma,P}. \tag{1384}$$

But every subset P of $\{1,2,\ldots,n\}$ satisfies

$$\begin{aligned}
&\underbrace{\sum_{\sigma\in S_n}}_{\substack{=\sum_{Q\subseteq\{1,2,\ldots,n\}}\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}}\\ \text{(because for every }\sigma\in S_n\text{, the set}\\ \sigma(P)\text{ is a subset of }\{1,2,\ldots,n\})}} c_{\sigma,P} \\
&= \sum_{Q\subseteq\{1,2,\ldots,n\}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} c_{\sigma,P} = \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} c_{\sigma,P} + \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|\neq|P|}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} c_{\sigma,P}}_{\substack{=0\\ \text{(by (1382))}}} \\
&\qquad \left(\begin{array}{c}\text{since every subset } Q \text{ of } \{1,2,\ldots,n\} \text{ satisfies}\\ \text{either } |Q|=|P| \text{ or } |Q|\neq|P| \text{ (but not both)}\end{array}\right) \\
&= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} c_{\sigma,P} + \underbrace{\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|\neq|P|}} 0}_{=0} \\
&= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(P)=Q}} c_{\sigma,P}}_{\substack{=(-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right)\\ \text{(by (1383))}}} \\
&= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right).
\end{aligned}$$

Hence, (1384) becomes

$$\begin{aligned}
\det(A+B) &= \sum_{P\subseteq\{1,2,\ldots,n\}} \underbrace{\sum_{\sigma\in S_n} c_{\sigma,P}}_{=\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right)} \\
&= \sum_{P\subseteq\{1,2,\ldots,n\}} \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |Q|=|P|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} B\right).
\end{aligned}$$

This proves Theorem 6.160. □

Solution to Exercise 6.48. Exercise 6.48 is solved, since Theorem 6.160 is proven. □

7.117. Solution to Exercise 6.49

Proof of Lemma 6.163. The set P is a subset of $\{1,2,\ldots,n\}$, and thus is finite (since $\{1,2,\ldots,n\}$ is finite). Hence, $|P| \in \mathbb{N}$.

Define an element $k \in \mathbb{N}$ by $k = |P|$. Thus, $k = |P| = |Q|$.

The list $w(P)$ is the list of all elements of P in increasing order (with no repetitions), and thus has k entries (since $|P| = k$). Thus, write this list $w(P)$ in the form $w(P) = (p_1, p_2, \ldots, p_k)$. Similarly, write the list $w(Q)$ in the form $w(Q) = (q_1, q_2, \ldots, q_k)$.

From $w(P) = (p_1, p_2, \ldots, p_k)$ and $w(Q) = (q_1, q_2, \ldots, q_k)$, we obtain

$$\begin{aligned} \operatorname{sub}_{w(P)}^{w(Q)} D &= \operatorname{sub}_{(p_1,p_2,\ldots,p_k)}^{(q_1,q_2,\ldots,q_k)} D = \operatorname{sub}_{p_1,p_2,\ldots,p_k}^{q_1,q_2,\ldots,q_k} D \\ &= \left(d_{p_x} \delta_{p_x,q_y}\right)_{1\leq x\leq k,\ 1\leq y\leq k} \end{aligned} \tag{1385}$$

(by the definition of $\operatorname{sub}_{p_1,p_2,\ldots,p_k}^{q_1,q_2,\ldots,q_k} D$, since $D = \left(d_i \delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Thus, in particular, we see that $\operatorname{sub}_{w(P)}^{w(Q)} D$ is a $k \times k$-matrix.

Recall that $(p_1, p_2, \ldots, p_k) = w(P)$ is a list of all elements of P (with no repetitions). Hence, $(p_1, p_2, \ldots, p_k)$ is a list with no repetitions, and satisfies $\{p_1, p_2, \ldots, p_k\} = P$. Now, every $(x,y) \in \{1,2,\ldots,k\}^2$ satisfies

$$\delta_{p_x,p_y} = \delta_{x,y} \tag{1386}$$

[612].

Recall again that $(p_1, p_2, \ldots, p_k)$ is a list of all elements of P (with no repetitions). Thus, the elements of P are $p_1, p_2, \ldots, p_k$ (with no repetitions). Hence,

$$\prod_{i\in P} d_i = d_{p_1} d_{p_2} \cdots d_{p_k} = \prod_{i=1}^{k} d_{p_i}. \tag{1387}$$

We are in one of the following two cases:

Case 1: We have $P \neq Q$.

Case 2: We have $P = Q$.

Let us consider Case 1 first. In this case, we have $P \neq Q$. Hence, $\delta_{P,Q} = 0$. On the other hand, there exists some $p \in P$ such that $p \notin Q$ [613]. Consider this p.

---

[612] *Proof of (1386):* Let $(x,y) \in \{1,2,\ldots,k\}^2$. Then, the integers $p_1, p_2, \ldots, p_k$ are pairwise distinct (since $(p_1, p_2, \ldots, p_k)$ is a list with no repetitions). Hence, $p_x = p_y$ holds if and only if $x = y$. Thus, $\begin{cases} 1, & \text{if } p_x = p_y; \\ 0, & \text{if } p_x \neq p_y \end{cases} = \begin{cases} 1, & \text{if } x = y; \\ 0, & \text{if } x \neq y \end{cases}$. In view of $\delta_{p_x,p_y} = \begin{cases} 1, & \text{if } p_x = p_y; \\ 0, & \text{if } p_x \neq p_y \end{cases}$ and $\delta_{x,y} = \begin{cases} 1, & \text{if } x = y; \\ 0, & \text{if } x \neq y \end{cases}$, this rewrites as $\delta_{p_x,p_y} = \delta_{x,y}$. This proves (1386).

[613] *Proof.* Assume the contrary (for the sake of contradiction). Hence, every $p \in P$ satisfies $p \in Q$. In other words, $P \subseteq Q$. But $|P| = |Q| \geq |Q|$.

Now, $p \in P = \{p_1, p_2, \ldots, p_k\}$. In other words, $p = p_u$ for some $u \in \{1, 2, \ldots, k\}$. Consider this u. Every $y \in \{1, 2, \ldots, k\}$ satisfies

$$\delta_{p_u, q_y} = 0 \tag{1388}$$

[614]. Now,

$$\begin{aligned} &\left(\text{the } u\text{-th row of the matrix } \underbrace{\operatorname{sub}_{w(P)}^{w(Q)} D}_{=\left(d_{p_x}\delta_{p_x,q_y}\right)_{1\leq x\leq k,\ 1\leq y\leq k}} \right) \\ &= \left(\text{the } u\text{-th row of the matrix } \left(d_{p_x}\delta_{p_x,q_y}\right)_{1\leq x\leq k,\ 1\leq y\leq k} \right) \\ &= \left(d_{p_u} \quad \underbrace{\delta_{p_u,q_y}}_{\substack{=0 \\ \text{(by (1388))}}} \right)_{1\leq x\leq 1,\ 1\leq y\leq k} = \left(\underbrace{d_{p_u} 0}_{=0} \right)_{1\leq x\leq 1,\ 1\leq y\leq k} = (0)_{1\leq x\leq 1,\ 1\leq y\leq k}. \end{aligned}$$

In other words, the u-th row of the matrix $\operatorname{sub}_{w(P)}^{w(Q)} D$ consists of zeroes. Therefore, Exercise 6.7 **(c)** (applied to k and $\operatorname{sub}_{w(P)}^{w(Q)} D$ instead of n and A) yields $\det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = 0$. Comparing this with $\underbrace{\delta_{P,Q}}_{=0} \prod_{i\in P} d_i = 0$, we obtain $\det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = \delta_{P,Q} \prod_{i\in P} d_i$. Hence, Lemma 6.163 is proven in Case 1.

Let us now consider Case 2. In this case, we have $P = Q$. Hence, $\delta_{P,Q} = 1$. On the other hand, from $w(Q) = (q_1, q_2, \ldots, q_k)$, we obtain

$$(q_1, q_2, \ldots, q_k) = w\left(\underbrace{Q}_{=P}\right) = w(P) = (p_1, p_2, \ldots, p_k).$$

In other words, $q_y = p_y$ for every $y \in \{1, 2, \ldots, k\}$. Thus, every $y \in \{1, 2, \ldots, k\}$ satisfies

$$\begin{aligned} \delta_{p_x,q_y} &= \delta_{p_x,p_y} && \left(\text{since } q_y = p_y\right) \\ &= \delta_{x,y} && \text{(by (1386))}. \end{aligned} \tag{1389}$$

But if X is a finite set, then every subset of X that has size $\geq |X|$ must be X itself. In other words, if X is a finite set, and if Y is a subset of X satisfying $|Y| \geq |X|$, then $Y = X$. Applying this to $X = Q$ and $Y = P$, we conclude that $P = Q$ (since P is a subset of Q, and since Q is a finite set). This contradicts $P \neq Q$. This contradiction shows that our assumption was wrong; qed.

[614] *Proof of (1388):* Let $y \in \{1, 2, \ldots, k\}$.

We know that $(q_1, q_2, \ldots, q_k) = w(Q)$ is a list of all elements of Q. Hence, $\{q_1, q_2, \ldots, q_k\} = Q$. Now, $y \in \{1, 2, \ldots, k\}$, so that $q_y \in \{q_1, q_2, \ldots, q_k\} = Q$. If we had $p_u = q_y$, then we would have $p = p_u = q_y \in Q$, which would contradict $p \notin Q$. Thus, we cannot have $p_u = q_y$. In other words, we have $p_u \neq q_y$. Hence, $\delta_{p_u,q_y} = 0$. This proves (1388).

Now, (1385) becomes

$$\operatorname{sub}_{w(P)}^{w(Q)} D = \left(d_{p_x} \underbrace{\delta_{p_x,q_y}}_{\substack{=\delta_{x,y} \\ \text{(by (1389))}}} \right)_{1\leq x\leq k,\ 1\leq y\leq k} = \left(d_{p_x} \delta_{x,y} \right)_{1\leq x\leq k,\ 1\leq y\leq k} = \left(d_{p_i} \delta_{i,j} \right)_{1\leq i\leq k,\ 1\leq j\leq k}$$

(here, we have renamed the index (x,y) as (i,j)).

But we have $d_{p_i}\delta_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,k\}^2$ satisfying $i < j$ [615]. Hence, Exercise 6.3 (applied to k, $\operatorname{sub}_{w(P)}^{w(Q)} D$ and $d_{p_i}\delta_{i,j}$ instead of n, A and $a_{i,j}$) yields $\det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = \left(d_{p_1}\delta_{1,1}\right)\left(d_{p_2}\delta_{2,2}\right)\cdots\left(d_{p_k}\delta_{k,k}\right)$ (since $\operatorname{sub}_{w(P)}^{w(Q)} D = \left(d_{p_i}\delta_{i,j}\right)_{1\leq i\leq k,\ 1\leq j\leq k}$). Thus,

$$\det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = \left(d_{p_1}\delta_{1,1}\right)\left(d_{p_2}\delta_{2,2}\right)\cdots\left(d_{p_k}\delta_{k,k}\right) = \prod_{i=1}^{k} \left(d_{p_i} \underbrace{\delta_{i,i}}_{\substack{=1 \\ \text{(since } i=i)}} \right)$$
$$= \prod_{i=1}^{k} d_{p_i} = \prod_{i\in P} d_i \qquad \text{(by (1387))}.$$

Comparing this with $\underbrace{\delta_{P,Q}}_{=1} \prod_{i\in P} d_i = \prod_{i\in P} d_i$, we obtain $\det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = \delta_{P,Q} \prod_{i\in P} d_i$. Hence, Lemma 6.163 is proven in Case 2.

We now have proven Lemma 6.163 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Lemma 6.163 always holds. □

Proof of Corollary 6.162. We start by making some general observations:

For any subset I of $\{1,2,\ldots,n\}$, we let $\widetilde{I}$ denote the complement $\{1,2,\ldots,n\} \setminus I$ of I. Then, any two subsets P and Q of $\{1,2,\ldots,n\}$ satisfy

$$\delta_{\widetilde{P},\widetilde{Q}} = \delta_{P,Q} \tag{1390}$$

[616].

[615] *Proof.* Let $(i,j) \in \{1,2,\ldots,k\}^2$ be such that $i < j$. Thus, $i \neq j$ (since $i < j$), so that $\delta_{i,j} = 0$. Thus, $d_{p_i} \underbrace{\delta_{i,j}}_{=0} = 0$, qed.

[616] *Proof of (1390):* Let P and Q be two subsets of $\{1,2,\ldots,n\}$. If $P = Q$, then $\widetilde{P} = \widetilde{Q}$ and thus $\delta_{\widetilde{P},\widetilde{Q}} = 1 = \delta_{P,Q}$ (since $P = Q$ yields $\delta_{P,Q} = 1$). Therefore, if $P = Q$, then (1390) holds. Hence, for the rest of the proof of (1390), we WLOG assume that we don't have $P = Q$. In other words, we have $P \neq Q$. Thus, $\delta_{P,Q} = 0$.

Furthermore, if P and Q are two subsets of $\{1,2,\ldots,n\}$ satisfying $|P| = |Q|$, then

$$\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} D\right) = \delta_{P,Q} \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i \qquad (1391)$$

[617].

Now, the definition of $\widetilde{P}$ yields $\widetilde{P} = \{1,2,\ldots,n\}\setminus P$. Hence, $\widetilde{P}$ is again a subset of $\{1,2,\ldots,n\}$. The definition of $\widetilde{\widetilde{P}}$ yields

$$\widetilde{\widetilde{P}} = \{1,2,\ldots,n\}\setminus \underbrace{\widetilde{P}}_{=\{1,2,\ldots,n\}\setminus P} = \{1,2,\ldots,n\}\setminus(\{1,2,\ldots,n\}\setminus P) = P$$

(since $P\subseteq\{1,2,\ldots,n\}$). Similarly, $\widetilde{\widetilde{Q}} = Q$. If we had $\widetilde{P} = \widetilde{Q}$, then we would have $\widetilde{\widetilde{P}} = \widetilde{\widetilde{Q}}$, which would contradict $\widetilde{\widetilde{P}} = P \neq Q = \widetilde{\widetilde{Q}}$. Hence, we cannot have $\widetilde{P} = \widetilde{Q}$. Thus, we have $\widetilde{P}\neq\widetilde{Q}$, so that $\delta_{\widetilde{P},\widetilde{Q}} = 0$. Compared with $\delta_{P,Q} = 0$, this yields $\delta_{\widetilde{P},\widetilde{Q}} = \delta_{P,Q}$. This proves (1390).

[617] *Proof of (1391):* Let P and Q be two subsets of $\{1,2,\ldots,n\}$ satisfying $|P| = |Q|$.

The definition of $\widetilde{P}$ yields $\widetilde{P} = \{1,2,\ldots,n\}\setminus P \subseteq \{1,2,\ldots,n\}$. In other words, $\widetilde{P}$ is a subset of $\{1,2,\ldots,n\}$. The same argument (applied to Q instead of P) shows that $\widetilde{Q}$ is a subset of $\{1,2,\ldots,n\}$.

We have

$$\left|\underbrace{\widetilde{P}}_{=\{1,2,\ldots,n\}\setminus P}\right| = |\{1,2,\ldots,n\}\setminus P| = \underbrace{|\{1,2,\ldots,n\}|}_{=n} - |P| \qquad (\text{since } P\subseteq\{1,2,\ldots,n\})$$
$$= n - |P|.$$

The same argument (applied to Q instead of P) shows that $\left|\widetilde{Q}\right| = n - |Q|$. Now, $\left|\widetilde{P}\right| = n - \underbrace{|P|}_{=|Q|} = n - |Q| = \left|\widetilde{Q}\right|$. Thus, Lemma 6.163 (applied to $\widetilde{P}$ and $\widetilde{Q}$ instead of P and Q) yields

$$\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} D\right) = \underbrace{\delta_{\widetilde{P},\widetilde{Q}}}_{\substack{=\delta_{P,Q}\\ (\text{by } (1390))}} \underbrace{\prod_{i\in\widetilde{P}}}_{\substack{=\prod_{i\in\{1,2,\ldots,n\}\setminus P}\\ (\text{since } \widetilde{P}=\{1,2,\ldots,n\}\setminus P)}} d_i = \delta_{P,Q}\prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i.$$

This proves (1391).

Now, every subset P of $\{1,2,\ldots,n\}$ satisfies

$$\begin{aligned}
&\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \underbrace{\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} D\right)}_{\substack{=\delta_{P,Q}\prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i\\ \text{(by (1391))}}} \\
&= \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \delta_{P,Q} \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i \\
&= \underbrace{(-1)^{\sum P+\sum P}}_{\substack{=1\\ \text{(since } \sum P+\sum P=2\sum P \text{ is even)}}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \underbrace{\delta_{P,P}}_{=1} \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i \\
&\qquad + \sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|;\\ Q\neq P}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \underbrace{\delta_{P,Q}}_{\substack{=0\\ \text{(since } P\neq Q\\ \text{(since } Q\neq P))}} \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i \\
&\qquad \left(\begin{array}{c}\text{here, we have split off the addend for } Q=P \text{ from the sum}\\ \text{(since } P \text{ is a subset } Q \text{ of } \{1,2,\ldots,n\} \text{ satisfying } |P|=|Q|\text{)}\end{array}\right) \\
&= \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i \\
&\qquad + \underbrace{\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|;\\ Q\neq P}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) 0 \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i}_{=0} \\
&= \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i .
\end{aligned} \tag{1392}$$

Now, Theorem 6.160 (applied to $B=D$) yields

$$\begin{aligned}
&\det(A+D) \\
&= \sum_{P\subseteq\{1,2,\ldots,n\}} \underbrace{\sum_{\substack{Q\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})} D\right)}_{\substack{=\det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right)\prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i\\ \text{(by (1392))}}} \\
&= \sum_{P\subseteq\{1,2,\ldots,n\}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \prod_{i\in\{1,2,\ldots,n\}\setminus P} d_i .
\end{aligned}$$

This proves Corollary 6.162. □

Proof of Corollary 6.164. For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$. Then, $I_n = \left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ (by the definition of I_n). Now,

$$x \underbrace{I_n}_{=\left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}} = x\left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = \left(x\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$$

(by the definition of $x\left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$). Hence, Corollary 6.162 (applied to $d_i = x$ and $D = xI_n$) yields

$$\begin{aligned} \det\left(A + xI_n\right) &= \sum_{P\subseteq\{1,2,\ldots,n\}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \underbrace{\prod_{i\in\{1,2,\ldots,n\}\setminus P} x}_{\substack{=x^{|\{1,2,\ldots,n\}\setminus P|}=x^{|\{1,2,\ldots,n\}|-|P|} \\ \text{(since } |\{1,2,\ldots,n\}\setminus P|=|\{1,2,\ldots,n\}|-|P| \\ \text{(since } P\subseteq\{1,2,\ldots,n\}))}} \\ &= \sum_{P\subseteq\{1,2,\ldots,n\}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \underbrace{x^{|\{1,2,\ldots,n\}|-|P|}}_{\substack{=x^{n-|P|} \\ \text{(since } |\{1,2,\ldots,n\}|=n)}} \\ &= \sum_{P\subseteq\{1,2,\ldots,n\}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) x^{n-|P|}. \end{aligned}$$

This proves (474).

Furthermore, every subset P of $\{1,2,\ldots,n\}$ satisfies $|P| \in \{0,1,\ldots,n\}$ [618].

[618] *Proof.* Let P be a subset of $\{1,2,\ldots,n\}$. Then, P is a finite set (since $\{1,2,\ldots,n\}$ is a finite set), so that $|P| \in \mathbb{N}$. Also, P is a subset of $\{1,2,\ldots,n\}$, and thus we have $|P| \leq |\{1,2,\ldots,n\}| = n$. Combined with $|P| \in \mathbb{N}$, this yields $|P| \in \{0,1,\ldots,n\}$. Qed.

Now,

$$\det(A + xI_n) = \underbrace{\sum_{P \subseteq \{1,2,\ldots,n\}}}_{\substack{= \sum\limits_{k \in \{0,1,\ldots,n\}} \sum\limits_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = k}} \\ \text{(since every subset } P \text{ of } \{1,2,\ldots,n\} \\ \text{satisfies } |P| \in \{0,1,\ldots,n\})}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) x^{n-|P|}$$

$$= \underbrace{\sum_{k \in \{0,1,\ldots,n\}}}_{= \sum\limits_{k=0}^{n}} \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \underbrace{x^{n-|P|}}_{\substack{= x^{n-k} \\ \text{(since } |P| = k)}}$$

$$= \sum_{k=0}^{n} \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) x^{n-k}$$

$$= \sum_{k=0}^{n} \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \underbrace{x^{n-(n-k)}}_{\substack{= x^{k} \\ \text{(since } n-(n-k) = k)}}$$

(here, we have substituted $n - k$ for k in the outer sum)

$$= \sum_{k=0}^{n} \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) x^{k}$$

$$= \sum_{k=0}^{n} \left(\sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = n-k}} \det\left(\operatorname{sub}_{w(P)}^{w(P)} A\right) \right) x^{k}.$$

This proves (475). The proof of Corollary 6.164 is now complete. □

Solution to Exercise 6.49. We have proven Lemma 6.163, Corollary 6.162 and Corollary 6.164. Exercise 6.49 is thus solved. □

7.118. Solution to Exercise 6.50

Let us first prepare some auxiliary results and notations. Throughout Section 7.118, we shall be using the following conventions:

- For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.
- For every $n \in \mathbb{N}$, the sign $\sum\limits_{I \subseteq [n]}$ shall mean $\sum\limits_{I \in \mathcal{P}([n])}$, where $\mathcal{P}([n])$ denotes the powerset of $[n]$.

- We shall use the Iverson bracket notation introduced in Definition 3.48.
- If $a_p, a_{p+1}, \ldots, a_q$ are some elements of a noncommutative ring $\mathbb{L}$ (with p and q being integers satisfying $p \leq q+1$), then $\prod_{i=p}^{q} a_i$ is defined to be the product $a_p a_{p+1} \cdots a_q \in \mathbb{L}$ [619]. This definition of $\prod_{i=p}^{q} a_i$ extends the classical definition of $\prod_{i=p}^{q} a_i$ in the case when $a_p, a_{p+1}, \ldots, a_q$ are elements of a commutative ring.

  (Notice that we have thus defined products of the form $\prod_{i=p}^{q} a_i$ only. In contrast, products of the form $\prod_{i \in I} a_i$ for an arbitrary finite set I are **not** defined in a noncommutative ring, because it is not clear in what order their factors are to be multiplied.)

A few elementary remarks about noncommutative rings will be useful:

- If p and q are integers satisfying $p \leq q$, and if $a_p, a_{p+1}, \ldots, a_q$ are some elements of a noncommutative ring $\mathbb{L}$, then

  $$\prod_{i=p}^{q} a_i = \left(\prod_{i=p}^{q-1} a_i \right) a_q.$$

  In other words, we can always split off the factor for $i = q$ from a product of the form $\prod_{i=p}^{q} a_i$ as long as we have $p \leq q$.

- When $\mathbb{L}$ is a noncommutative ring, the following two properties of the $\sum$ sign hold[620]:

[619] Of course, in order for this definition to be fully rigorous, we need to specify how the product $a_p a_{p+1} \cdots a_q$ is defined. There are two possible definitions for this product; both of them proceed by recursion on $q - p$ (with the recursion base being the case when $q - p = -1$).

- The first definition sets $a_p a_{p+1} \cdots a_q$ to be 1 (the unity of $\mathbb{L}$) when $q - p = -1$ (this is the case of an empty product), and sets $a_p a_{p+1} \cdots a_q$ to be $\left(a_p a_{p+1} \cdots a_{q-1}\right) a_q$ when $q - p \geq 0$.
- The second definition sets $a_p a_{p+1} \cdots a_q$ to be 1 (the unity of $\mathbb{L}$) when $q - p = -1$ (this is the case of an empty product), and sets $a_p a_{p+1} \cdots a_q$ to be $a_p \left(a_{p+1} a_{p+2} \cdots a_q\right)$ when $q - p \geq 0$.

Fortunately, the two definitions define exactly the same notion of product; this is not hard to prove (it follows from a fact called "general associativity", which is analogous to Proposition 2.90 but involves multiplication of elements of $\mathbb{L}$ instead of composition of maps), but it is not entirely obvious.

[620] These are analogues of the "Factoring out" property that we have seen in Section 1.4.

– **Factoring out on the left:** Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{L}$. Also, let λ be an element of $\mathbb{L}$. Then,

$$\sum_{s\in S} \lambda a_s = \lambda \sum_{s\in S} a_s. \qquad (1393)$$

– **Factoring out on the right:** Let S be a finite set. For every $s \in S$, let a_s be an element of $\mathbb{L}$. Also, let λ be an element of $\mathbb{L}$. Then,

$$\sum_{s\in S} a_s \lambda = \left(\sum_{s\in S} a_s\right) \lambda. \qquad (1394)$$

Next, we state a generalization of Lemma 6.22 to the case of a noncommutative ring $\mathbb{K}$:

Lemma 7.220. Let $\mathbb{K}$ be a noncommutative ring.

Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. For every $i \in [n]$, let $p_{i,1}, p_{i,2}, \ldots, p_{i,m}$ be m elements of $\mathbb{K}$. Then,

$$\prod_{i=1}^{n} \sum_{k=1}^{m} p_{i,k} = \sum_{\kappa:[n]\to[m]} \prod_{i=1}^{n} p_{i,\kappa(i)}.$$

Proof of Lemma 7.220. The proof of Lemma 6.22 given above (which includes the proof of Lemma 6.20 in the solution to Exercise 6.9) can be reused as a proof of Lemma 7.220, provided that we replace our proof of Lemma 7.161 (which made use of the commutativity of $\mathbb{K}$) by the following alternative proof (which does not require $\mathbb{K}$ to be commutative):

Second proof of Lemma 7.161. The equality (27) remains valid if $\mathbb{A}$ is replaced by $\mathbb{K}$ throughout it. It is in this variant that it will be used in the following proof.

For every $\lambda \in \mathbb{K}$, we have

$$\begin{aligned}
\sum_{y\in Y} \lambda r_y &= \sum_{s\in Y} \lambda r_s && \text{(here, we renamed the summation index } y \text{ as } s\text{)} \\
&= \lambda \sum_{s\in Y} r_s && \left(\begin{array}{c}\text{by (1393) (applied to } Y,\ r_s \text{ and } \mathbb{K} \\ \text{instead of } S,\ a_s \text{ and } \mathbb{L}\text{)}\end{array}\right) \\
&= \lambda \sum_{y\in Y} r_y && \text{(here, we renamed the summation index } s \text{ as } y\text{)}. \qquad (1395)
\end{aligned}$$

For every $\lambda \in \mathbb{K}$, we have

$$\begin{aligned}
\sum_{x\in X} q_x \lambda &= \sum_{s\in X} q_s \lambda && \text{(here, we renamed the summation index } x \text{ as } s\text{)} \\
&= \left(\sum_{s\in X} q_s\right)\lambda && \left(\begin{array}{c}\text{by (1394) (applied to } X\text{, } q_s \text{ and } \mathbb{K} \\ \text{instead of } S\text{, } a_s \text{ and } \mathbb{L}\text{)}\end{array}\right) \\
&= \left(\sum_{x\in X} q_x\right)\lambda && \text{(here, we renamed the summation index } s \text{ as } x\text{)}.
\end{aligned} \tag{1396}$$

From (27) (applied to $\mathbb{K}$ and $q_x r_y$ instead of $\mathbb{A}$ and $a_{(x,y)}$), we obtain

$$\sum_{x\in X}\sum_{y\in Y} q_x r_y = \sum_{(x,y)\in X\times Y} q_x r_y = \sum_{y\in Y}\sum_{x\in X} q_x r_y.$$

Hence,

$$\begin{aligned}
\sum_{(x,y)\in X\times Y} q_x r_y &= \sum_{x\in X} \underbrace{\sum_{y\in Y} q_x r_y}_{\substack{=q_x \sum_{y\in Y} r_y \\ \text{(by (1395) (applied to } \lambda=q_x\text{))}}} = \sum_{x\in X} q_x \sum_{y\in Y} r_y \\
&= \left(\sum_{x\in X} q_x\right)\left(\sum_{y\in Y} r_y\right)
\end{aligned}$$

(by (1396) (applied to $\lambda = \sum_{y\in Y} r_y$)). This proves Lemma 7.161. ☐

Thus, Lemma 7.220 is proven. ☐

Now, let us prove some further lemmas:

Lemma 7.221. Let G be a finite set. Let I be a subset of G. Let $m \in \mathbb{N}$. Let $f : [m] \to G$ be any map. Then,

$$\prod_{i=1}^{m} [f(i) \in I] = [f([m]) \subseteq I].$$

Proof of Lemma 7.221. We are in one of the following two cases:

Case 1: We have $f([m]) \subseteq I$.

Case 2: We don't have $f([m]) \subseteq I$.

Let us first consider Case 1. In this case, we have $f([m]) \subseteq I$. Thus, $[f([m]) \subseteq I] = 1$. But each $i \in \{1, 2, \ldots, m\}$ satisfies $[f(i) \in I] = 1$ (since $f\left(\underbrace{i}_{\in[m]}\right) \in f([m]) \subseteq I$).

Hence, $\prod_{i=1}^{m} \underbrace{[f(i) \in I]}_{=1} = \prod_{i=1}^{m} 1 = 1$. Comparing this with $[f([m]) \subseteq I] = 1$, we obtain $\prod_{i=1}^{m} [f(i) \in I] = [f([m]) \subseteq I]$. Hence, Lemma 7.221 is proven in Case 1.

Let us now consider Case 2. In this case, we don't have $f([m]) \subseteq I$. In other words, not every $p \in [m]$ satisfies $f(p) \in I$. Thus, there exists some $p \in [m]$ such that $f(p) \notin I$. Consider this p. We have $[f(p) \in I] = 0$ (since $f(p) \notin I$). Thus, at least one factor of the product $\prod_{i=1}^{m} [f(i) \in I]$ equals 0 (namely, the factor for $i = p$ is $[f(p) \in I] = 0$). Consequently, the whole product $\prod_{i=1}^{m} [f(i) \in I]$ equals 0 (because if at least one factor of a product equals 0, then the whole product equals 0). In other words, $\prod_{i=1}^{m} [f(i) \in I] = 0$. Comparing this with $[f([m]) \subseteq I] = 0$ (which holds because we don't have $f([m]) \subseteq I$), we obtain $\prod_{i=1}^{m} [f(i) \in I] = [f([m]) \subseteq I]$. Hence, Lemma 7.221 is proven in Case 2.

We have now proven Lemma 7.221 in both Cases 1 and 2. This completes the proof of Lemma 7.221. □

Lemma 7.222. Let $\mathbb{L}$ be a noncommutative ring. Let $n \in \mathbb{N}$. Let $v_1, v_2, \ldots, v_n$ be n elements of $\mathbb{L}$. Let I be a subset of $[n]$. Then,

$$\left(\sum_{i \in I} v_i\right)^m = \sum_{\substack{f:[m]\to[n]; \\ f([m])\subseteq I}} v_{f(1)} v_{f(2)} \cdots v_{f(m)}.$$

Proof of Lemma 7.222. First, observe that

$$\sum_{k=1}^{n}[k\in I]v_k = \underbrace{\sum_{i=1}^{n}}_{\substack{=\sum_{i\in\{1,2,\ldots,n\}}=\sum_{i\in[n]} \\ (\text{since } \{1,2,\ldots,n\}=[n])}} [i\in I]v_i \qquad \left(\begin{array}{c}\text{here, we have renamed the} \\ \text{summation index } k \text{ as } i\end{array}\right)$$
$$= \sum_{i\in[n]}[i\in I]v_i$$
$$= \underbrace{\sum_{\substack{i\in[n];\\ i\in I}}}_{\substack{=\sum_{i\in I} \\ (\text{since } I\subseteq[n])}} \underbrace{[i\in I]}_{\substack{=1 \\ (\text{since } i\in I \text{ is true})}} v_i + \sum_{\substack{i\in[n];\\ i\notin I}} \underbrace{[i\in I]}_{\substack{=0 \\ (\text{since } i\in I \text{ is false} \\ (\text{because } i\notin I))}} v_i$$
$$\left(\text{since each } i\in[n] \text{ satisfies either } i\in I \text{ or } i\notin I \text{ (but not both)}\right)$$
$$= \sum_{i\in I}1v_i + \underbrace{\sum_{\substack{i\in[n];\\ i\notin I}}0v_i}_{=0} = \sum_{i\in I}1v_i = \sum_{i\in I}v_i. \tag{1397}$$

But $p^m = \prod_{i=1}^{m} p$ for each $p\in\mathbb{L}$. Applying this to $p = \sum_{k=1}^{n}[k\in I]v_k$, we find

$$\left(\sum_{k=1}^{n}[k\in I]v_k\right)^m = \prod_{i=1}^{m}\sum_{k=1}^{n}[k\in I]v_k = \sum_{\kappa:[m]\to[n]}\prod_{i=1}^{m}\left([\kappa(i)\in I]v_{\kappa(i)}\right)$$

(by Lemma 7.220 (applied to $\mathbb{L}$, m, n and $[k\in I]v_k$ instead of $\mathbb{K}$, n, m and $p_{i,k}$)). Now,

$$\left(\underbrace{\sum_{i\in I}v_i}_{\substack{=\sum_{k=1}^{n}[k\in I]v_k \\ (\text{by } (1397))}}\right)^m = \left(\sum_{k=1}^{n}[k\in I]v_k\right)^m$$
$$= \sum_{\kappa:[m]\to[n]}\prod_{i=1}^{m}\left([\kappa(i)\in I]v_{\kappa(i)}\right)$$
$$= \sum_{f:[m]\to[n]}\prod_{i=1}^{m}\left([f(i)\in I]v_{f(i)}\right) \tag{1398}$$

(here, we have renamed the summation index κ as f).

Now, fix any map $f : [m] \to [n]$. Then,

$$\prod_{i=1}^{m} \left([f(i) \in I] v_{f(i)} \right) = \left([f(1) \in I] v_{f(1)} \right) \left([f(2) \in I] v_{f(2)} \right) \cdots \left([f(m) \in I] v_{f(m)} \right). \tag{1399}$$

The factors $[f(1) \in I], [f(2) \in I], \ldots, [f(m) \in I]$ on the right hand side of this equality are integers, and therefore can be freely moved within the product (even though $\mathbb{L}$ is not necessarily commutative). In particular, we can move them to the front of the product. Thus, we find

$$\begin{aligned} & \left([f(1) \in I] v_{f(1)} \right) \left([f(2) \in I] v_{f(2)} \right) \cdots \left([f(m) \in I] v_{f(m)} \right) \\ &= \underbrace{\left([f(1) \in I] [f(2) \in I] \cdots [f(m) \in I] \right)}_{= \prod_{i=1}^{m} [f(i) \in I]} \left(v_{f(1)} v_{f(2)} \cdots v_{f(m)} \right) \\ &= \left(\prod_{i=1}^{m} [f(i) \in I] \right) \left(v_{f(1)} v_{f(2)} \cdots v_{f(m)} \right). \end{aligned}$$

Hence, (1399) rewrites as

$$\prod_{i=1}^{m} \left([f(i) \in I] v_{f(i)} \right) = \left(\prod_{i=1}^{m} [f(i) \in I] \right) \left(v_{f(1)} v_{f(2)} \cdots v_{f(m)} \right). \tag{1400}$$

But Lemma 7.221 (applied to $G = [n]$) yields

$$\prod_{i=1}^{m} [f(i) \in I] = [f([m]) \subseteq I]. \tag{1401}$$

Hence, (1400) becomes

$$\begin{aligned} \prod_{i=1}^{m} \left([f(i) \in I] v_{f(i)} \right) &= \underbrace{\left(\prod_{i=1}^{m} [f(i) \in I] \right)}_{\substack{= [f([m]) \subseteq I] \\ \text{(by (1401))}}} \left(v_{f(1)} v_{f(2)} \cdots v_{f(m)} \right) \\ &= [f([m]) \subseteq I] \left(v_{f(1)} v_{f(2)} \cdots v_{f(m)} \right). \end{aligned} \tag{1402}$$

Now, forget that we fixed f. We thus have proven (1402) for each map $f : [m] \to [n]$.

Now, (1398) becomes

$$\begin{aligned}
\left(\sum_{i\in I} v_i\right)^m &= \sum_{f:[m]\to[n]} \underbrace{\prod_{i=1}^{m}\left([f(i)\in I]\, v_{f(i)}\right)}_{\substack{=[f([m])\subseteq I]\left(v_{f(1)}v_{f(2)}\cdots v_{f(m)}\right) \\ \text{(by (1402))}}} \\
&= \sum_{f:[m]\to[n]} [f([m])\subseteq I]\left(v_{f(1)}v_{f(2)}\cdots v_{f(m)}\right) \\
&= \sum_{\substack{f:[m]\to[n];\\ f([m])\subseteq I}} \underbrace{[f([m])\subseteq I]}_{\substack{=1\\ \text{(since } f([m])\subseteq I)}}\left(v_{f(1)}v_{f(2)}\cdots v_{f(m)}\right) \\
&\qquad + \sum_{\substack{f:[m]\to[n];\\ \text{not } f([m])\subseteq I}} \underbrace{[f([m])\subseteq I]}_{\substack{=0\\ \text{(since we don't have } f([m])\subseteq I)}}\left(v_{f(1)}v_{f(2)}\cdots v_{f(m)}\right) \\
&= \sum_{\substack{f:[m]\to[n];\\ f([m])\subseteq I}} v_{f(1)}v_{f(2)}\cdots v_{f(m)} + \underbrace{\sum_{\substack{f:[m]\to[n];\\ \text{not } f([m])\subseteq I}} 0\left(v_{f(1)}v_{f(2)}\cdots v_{f(m)}\right)}_{=0} \\
&= \sum_{\substack{f:[m]\to[n];\\ f([m])\subseteq I}} v_{f(1)}v_{f(2)}\cdots v_{f(m)}.
\end{aligned}$$

This proves Lemma 7.222. □

Lemma 7.223. Let $n\in\mathbb{N}$. Let J be a subset of $[n]$. Then,

$$\sum_{\substack{I\subseteq[n];\\ J\subseteq I}} (-1)^{n-|I|} = [J=[n]].$$

Proof of Lemma 7.223. Here is a proof that makes use of Exercise 6.1. It is not the simplest possible proof, but it might be the shortest.[621]

We first make the following observations:

- If I is a subset of $[n]$ satisfying $J\subseteq I$, then
$$\prod_{i\in[n]\setminus I} [i\notin J] = 1 \tag{1403}$$
[622].

[621] We are working in the commutative ring $\mathbb{Z}$ in this proof. Hence, product signs such as $\prod_{i\in I}$ and $\prod_{i\in[n]\setminus I}$ make sense.

[622] *Proof of (1403):* Let I be a subset of $[n]$ satisfying $J\subseteq I$.

- If I is a subset of $[n]$ that does **not** satisfy $J \subseteq I$, then
$$\prod_{i \in [n] \setminus I} [i \notin J] = 0 \qquad (1404)$$
[623].

But Exercise 6.1 **(a)** (applied to $\mathbb{Z}$, 1 and $-[i \notin J]$ instead of $\mathbb{K}$, a_i and b_i) yields

$$\begin{aligned}
&\prod_{i=1}^{n} (1 + (-[i \notin J])) \\
&= \sum_{I \subseteq [n]} \underbrace{\left(\prod_{i \in I} 1\right)}_{=1} \underbrace{\left(\prod_{i \in [n] \setminus I} (-[i \notin J])\right)}_{=(-1)^{|[n] \setminus I|} \prod_{i \in [n] \setminus I} [i \notin J]} \\
&= \sum_{I \subseteq [n]} (-1)^{|[n] \setminus I|} \prod_{i \in [n] \setminus I} [i \notin J] \\
&= \sum_{\substack{I \subseteq [n]; \\ J \subseteq I}} \underbrace{(-1)^{|[n] \setminus I|}}_{\substack{=(-1)^{n-|I|} \\ (\text{since } |[n] \setminus I| = n - |I| \\ (\text{since } I \subseteq [n]))}} \underbrace{\prod_{i \in [n] \setminus I} [i \notin J]}_{\substack{=1 \\ (\text{by } (1403))}} + \sum_{\substack{I \subseteq [n]; \\ \text{not } J \subseteq I}} (-1)^{|[n] \setminus I|} \underbrace{\prod_{i \in [n] \setminus I} [i \notin J]}_{\substack{=0 \\ (\text{by } (1404))}} \\
&\qquad \left(\begin{array}{c} \text{since each subset } I \text{ of } [n] \text{ satisfies either } J \subseteq I \text{ or } (\text{not } J \subseteq I), \\ \text{but not both} \end{array}\right) \\
&= \sum_{\substack{I \subseteq [n]; \\ J \subseteq I}} (-1)^{n-|I|} + \underbrace{\sum_{\substack{I \subseteq [n]; \\ \text{not } J \subseteq I}} (-1)^{|[n] \setminus I|} 0}_{=0} = \sum_{\substack{I \subseteq [n]; \\ J \subseteq I}} (-1)^{n-|I|}. \qquad (1405)
\end{aligned}$$

On the other hand, it is easy to see that

$$\prod_{i=1}^{n} (1 + (-[i \notin J])) = [J = [n]] \qquad (1406)$$

From $J \subseteq I$, we obtain $[n] \setminus I \subseteq [n] \setminus J$. Thus, each $i \in [n] \setminus I$ satisfies $i \in [n] \setminus I \subseteq [n] \setminus J$ and therefore $i \notin J$. Hence, each $i \in [n] \setminus I$ satisfies $[i \notin J] = 1$. Therefore, $\prod_{i \in [n] \setminus I} \underbrace{[i \notin J]}_{=1} = \prod_{i \in [n] \setminus I} 1 = 1$. This proves (1403).

[623] *Proof of (1404):* Let I be a subset of $[n]$ that does **not** satisfy $J \subseteq I$.

We have $J \not\subseteq I$ (since we do not have $J \subseteq I$). Hence, there exists some $j \in J$ such that $j \notin I$. Consider this j. We have $[j \notin J] = 0$ (since $j \in J$). Combining $j \in J \subseteq [n]$ with $j \notin I$, we obtain $j \in [n] \setminus I$. Hence, at least one factor of the product $\prod_{i \in [n] \setminus I} [i \notin J]$ is 0 (namely, the factor for $i = j$ is $[j \notin J] = 0$). Thus, the whole product $\prod_{i \in [n] \setminus I} [i \notin J]$ must be 0 (because if a factor of a product is 0, then the whole product must be 0). This proves (1404).

[624]. Comparing this with (1405), we obtain

$$\sum_{\substack{I \subseteq [n]; \\ J \subseteq I}} (-1)^{n-|I|} = [J = [n]] .$$

This proves Lemma 7.223. □

Solution to Exercise 6.50. **(a)** Let $m \in \mathbb{N}$.

[624] *Proof of (1406):* We are in one of the following two cases:

Case 1: We have $J = [n]$.

Case 2: We don't have $J = [n]$.

Let us first consider Case 1. In this case, we have $J = [n]$. Thus, $[J = [n]] = 1$. But each $i \in \{1, 2, \ldots, n\}$ satisfies $i \in \{1, 2, \ldots, n\} = [n] = J$ and therefore $[i \notin J] = 0$. Hence,

$$\prod_{i=1}^{n} \left(1 + \left(- \underbrace{[i \notin J]}_{=0} \right) \right) = \prod_{i=1}^{n} \underbrace{(1 + (-0))}_{=1} = \prod_{i=1}^{n} 1 = 1.$$

Comparing this with $[J = [n]] = 1$, we obtain $\prod_{i=1}^{n} (1 + (-[i \notin J])) = [J = [n]]$. Hence, (1406) is proven in Case 1.

Let us now consider Case 2. In this case, we don't have $J = [n]$. Hence, $J \neq [n]$. Thus, J is a **proper** subset of $[n]$ (since $J \subseteq [n]$). Therefore, there exists some $j \in [n]$ such that $j \notin J$. Consider this j. We have $[j \notin J] = 1$ (since $j \notin J$). But $j \in [n] = \{1, 2, \ldots, n\}$. Thus, at least one factor of the product $\prod_{i=1}^{n} (1 + (-[i \notin J]))$ equals 0 (namely, the factor for $i = j$ is $1 + \left(- \underbrace{[j \notin J]}_{=1} \right) = 1 + (-1) = 0$). Consequently, the whole product $\prod_{i=1}^{n} (1 + (-[i \notin J]))$ equals 0 (because if at least one factor of a product equals 0, then the whole product equals 0). In other words, $\prod_{i=1}^{n} (1 + (-[i \notin J])) = 0$. Comparing this with $[J = [n]] = 0$ (which holds because we don't have $J = [n]$), we obtain $\prod_{i=1}^{n} (1 + (-[i \notin J])) = [J = [n]]$. Hence, (1406) is proven in Case 2.

We have now proven (1406) in both Cases 1 and 2. Thus, (1406) is proven.

We have

$$\begin{aligned}
&\sum_{I\subseteq[n]} (-1)^{n-|I|} \underbrace{\left(\sum_{i\in I} v_i\right)^m}_{\substack{=\sum_{\substack{f:[m]\to[n];\\ f([m])\subseteq I}} v_{f(1)}v_{f(2)}\cdots v_{f(m)} \\ \text{(by Lemma 7.222)}}} \\
&= \sum_{I\subseteq[n]} (-1)^{n-|I|} \sum_{\substack{f:[m]\to[n];\\ f([m])\subseteq I}} v_{f(1)}v_{f(2)}\cdots v_{f(m)} \\
&= \underbrace{\sum_{I\subseteq[n]} \sum_{\substack{f:[m]\to[n];\\ f([m])\subseteq I}}}_{=\sum_{f:[m]\to[n]} \sum_{\substack{I\subseteq[n];\\ f([m])\subseteq I}}} (-1)^{n-|I|} v_{f(1)}v_{f(2)}\cdots v_{f(m)} \\
&= \sum_{f:[m]\to[n]} \sum_{\substack{I\subseteq[n];\\ f([m])\subseteq I}} (-1)^{n-|I|} v_{f(1)}v_{f(2)}\cdots v_{f(m)} \\
&= \sum_{f:[m]\to[n]} \underbrace{\left(\sum_{\substack{I\subseteq[n];\\ f([m])\subseteq I}} (-1)^{n-|I|}\right)}_{\substack{=[f([m])=[n]] \\ \text{(by Lemma 7.223} \\ \text{(applied to } J=f([m])))}} v_{f(1)}v_{f(2)}\cdots v_{f(m)}
\end{aligned}$$

$$= \sum_{f:[m]\to[n]} \left[\underbrace{f([m]) = [n]}_{\iff (f \text{ is surjective})} \right] v_{f(1)} v_{f(2)} \cdots v_{f(m)}$$
$$= \sum_{f:[m]\to[n]} [f \text{ is surjective}]\, v_{f(1)} v_{f(2)} \cdots v_{f(m)}$$
$$= \sum_{\substack{f:[m]\to[n];\\ f \text{ is surjective}}} \underbrace{[f \text{ is surjective}]}_{\substack{=1 \\ (\text{since } f \text{ is surjective})}} v_{f(1)} v_{f(2)} \cdots v_{f(m)} + \sum_{\substack{f:[m]\to[n];\\ f \text{ is not surjective}}} \underbrace{[f \text{ is surjective}]}_{\substack{=0 \\ (\text{since } f \text{ is not surjective})}} v_{f(1)} v_{f(2)} \cdots v_{f(m)}$$
$$= \sum_{\substack{f:[m]\to[n];\\ f \text{ is surjective}}} v_{f(1)} v_{f(2)} \cdots v_{f(m)} + \underbrace{\sum_{\substack{f:[m]\to[n];\\ f \text{ is not surjective}}} 0 v_{f(1)} v_{f(2)} \cdots v_{f(m)}}_{=0}$$
$$= \sum_{\substack{f:[m]\to[n];\\ f \text{ is surjective}}} v_{f(1)} v_{f(2)} \cdots v_{f(m)}.$$

This solves Exercise 6.50 **(a)**.

(b) Let $m \in \{0, 1, \ldots, n-1\}$. Then, there exists no map $f : [m] \to [n]$ such that f is surjective[625]. Hence, the sum $\sum_{\substack{f:[m]\to[n];\\ f \text{ is surjective}}} v_{f(1)} v_{f(2)} \cdots v_{f(m)}$ is an empty sum. Therefore, this sum equals 0. In other words, $\sum_{\substack{f:[m]\to[n];\\ f \text{ is surjective}}} v_{f(1)} v_{f(2)} \cdots v_{f(m)} = 0$.

Now, Exercise 6.50 **(a)** yields

$$\sum_{I\subseteq[n]} (-1)^{n-|I|} \left(\sum_{i\in I} v_i \right)^m = \sum_{\substack{f:[m]\to[n];\\ f \text{ is surjective}}} v_{f(1)} v_{f(2)} \cdots v_{f(m)} = 0.$$

This solves Exercise 6.50 **(b)**.

(c) Clearly, $[n]$ and $[n]$ are two finite sets such that $|[n]| \leq |[n]|$. Hence, Lemma 1.4 (applied to $U = [n]$ and $V = [n]$) shows that if $f : [n] \to [n]$ is a map, then we have the following logical equivalence:

$$(f \text{ is surjective}) \iff (f \text{ is bijective}). \qquad (1407)$$

[625] *Proof.* Let $f : [m] \to [n]$ be a map such that f is surjective. Thus, $f([m]) = [n]$ (since f is surjective). But clearly, $|f([m])| \leq |[m]| = m$. Since $f([m]) = [n]$, this rewrites as $|[n]| \leq m$. Since $|[n]| = n$, this rewrites as $n \leq m$. But $m \leq n-1$ (since $m \in \{0, 1, \ldots, n-1\}$). Thus, $n \leq m \leq n-1 < n$. This is absurd.

Now, forget that we fixed f. We thus have found a contradiction for each map $f : [m] \to [n]$ such that f is surjective. Hence, there exists no map $f : [m] \to [n]$ such that f is surjective. Qed.

Hence, we have the following equality of summation signs:

$$\sum_{\substack{f:[n]\to[n];\\ f \text{ is surjective}}} = \sum_{\substack{f:[n]\to[n];\\ f \text{ is bijective}}} = \sum_{f \text{ is a permutation of } [n]} = \sum_{f\in S_n} \tag{1407}$$

(since S_n is the set of all permutations of $[n]$). Now, Exercise 6.50 **(a)** (applied to $m = n$) yields

$$\sum_{I\subseteq[n]} (-1)^{n-|I|} \left(\sum_{i\in I} v_i\right)^n = \underbrace{\sum_{\substack{f:[n]\to[n];\\ f \text{ is surjective}}}}_{\substack{=\sum_{f\in S_n}\\ \text{(by (1407))}}} v_{f(1)}v_{f(2)}\cdots v_{f(n)}$$
$$= \sum_{f\in S_n} v_{f(1)}v_{f(2)}\cdots v_{f(n)} = \sum_{\sigma\in S_n} v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)}$$

(here, we have renamed the summation index f as σ). This solves Exercise 6.50 **(c)**.

(d) We have

$$v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)} = v_1v_2\cdots v_n \tag{1408}$$

for each $\sigma \in S_n$ [626]. Now, Exercise 6.50 **(c)** yields

$$\sum_{I\subseteq[n]} (-1)^{n-|I|} \left(\sum_{i\in I} v_i\right)^n = \sum_{\sigma\in S_n} \underbrace{v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)}}_{\substack{=v_1v_2\cdots v_n\\ \text{(by (1408))}}} = \sum_{\sigma\in S_n} v_1v_2\cdots v_n$$
$$= \underbrace{|S_n|}_{=n!} v_1v_2\cdots v_n = n!\cdot v_1v_2\cdots v_n.$$

This solves Exercise 6.50 **(d)**. □

[626] *Proof of (1408):* Let $\sigma \in S_n$. Thus, σ is a permutation of the set $\{1,2,\ldots,n\}$ (since S_n is the set of all permutations of the set $\{1,2,\ldots,n\}$). In other words, σ is a bijection $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$. But

$$v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)} = \underbrace{\prod_{i=1}^{n}}_{=\prod_{i\in\{1,2,\ldots,n\}}} v_{\sigma(i)} = \prod_{i\in\{1,2,\ldots,n\}} v_{\sigma(i)} = \underbrace{\prod_{i\in\{1,2,\ldots,n\}}}_{=\prod_{i=1}^{n}} v_i$$
$$\left(\begin{array}{c}\text{here, we have substituted } i \text{ for } \sigma(i) \text{ in the product,}\\ \text{since } \sigma \text{ is a bijection } \{1,2,\ldots,n\}\to\{1,2,\ldots,n\}\end{array}\right)$$
$$= \prod_{i=1}^{n} v_i = v_1v_2\cdots v_n.$$

This proves (1408).

7.119. Solution to Exercise 6.51

Throughout Section 7.119, we shall use the same notations that we have used in Section 7.118. Also, for any set S, we let $\mathcal{P}(S)$ denote the powerset of S (that is, the set of all subsets of S).

Solution to Exercise 6.51. We have $n+1 \in [n+1]$ (since $n+1$ is a positive integer).

Also, $[n+1] \setminus \{n+1\} = [n]$ (for the same reason).

But Fact 1 from the solution to Exercise 6.1 (applied to $S = [n+1]$ and $s = n+1$) yields the following two facts:

- We have $\mathcal{P}([n+1] \setminus \{n+1\}) \subseteq \mathcal{P}([n+1])$.
- The map

$$\mathcal{P}([n+1] \setminus \{n+1\}) \to \mathcal{P}([n+1]) \setminus \mathcal{P}([n+1] \setminus \{n+1\}),$$
$$U \mapsto U \cup \{n+1\}$$

  is well-defined and a bijection.

Using $[n+1] \setminus \{n+1\} = [n]$, we can rewrite these two facts as follows:

- We have $\mathcal{P}([n]) \subseteq \mathcal{P}([n+1])$.
- The map

$$\mathcal{P}([n]) \to \mathcal{P}([n+1]) \setminus \mathcal{P}([n]),$$
$$U \mapsto U \cup \{n+1\}$$

  is well-defined and a bijection.

We now claim that

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(w + \sum_{i \in I} v_i \right)^m = \sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m \qquad (1409)$$

for every $m \in \{0, 1, \ldots, n\}$ and $w \in \mathbb{L}$.

[*Proof of (1409):* Let $m \in \{0, 1, \ldots, n\}$ and $w \in \mathbb{L}$.

We extend the n-tuple $(v_1, v_2, \ldots, v_n) \in \mathbb{L}^n$ to an $(n+1)$-tuple $(v_1, v_2, \ldots, v_{n+1}) \in \mathbb{L}^{n+1}$ by setting $v_{n+1} = w$. Thus, $v_1, v_2, \ldots, v_{n+1}$ are $n+1$ elements of $\mathbb{L}$. Moreover, $m \in \{0, 1, \ldots, n\} = \{0, 1, \ldots, (n+1)-1\}$. Hence, Exercise 6.50 **(b)** (applied to $n+1$ instead of n) yields

$$\sum_{I \subseteq [n+1]} (-1)^{(n+1)-|I|} \left(\sum_{i \in I} v_i \right)^m = 0.$$

Hence,

$$0 = \underbrace{\sum_{I \subseteq [n+1]}}_{= \sum_{I \in \mathcal{P}([n+1])}} (-1)^{(n+1)-|I|} \left(\sum_{i \in I} v_i \right)^m = \sum_{I \in \mathcal{P}([n+1])} (-1)^{(n+1)-|I|} \left(\sum_{i \in I} v_i \right)^m$$

$$= \underbrace{\sum_{\substack{I \in \mathcal{P}([n+1]); \\ I \in \mathcal{P}([n])}}}_{\substack{= \sum_{I \in \mathcal{P}([n])} \\ (\text{since } \mathcal{P}([n]) \subseteq \mathcal{P}([n+1]))}} \underbrace{(-1)^{(n+1)-|I|}}_{\substack{= -(-1)^{n-|I|} \\ (\text{since } (n+1)-|I| = (n-|I|)+1)}} \left(\sum_{i \in I} v_i \right)^m$$

$$+ \underbrace{\sum_{\substack{I \in \mathcal{P}([n+1]); \\ I \notin \mathcal{P}([n])}}}_{= \sum_{I \in \mathcal{P}([n+1]) \setminus \mathcal{P}([n])}} (-1)^{(n+1)-|I|} \left(\sum_{i \in I} v_i \right)^m$$

$$\left(\begin{array}{c} \text{since each } I \in \mathcal{P}([n+1]) \text{ satisfies either } I \in \mathcal{P}([n]) \\ \text{or } I \notin \mathcal{P}([n]) \text{ (but not both)} \end{array} \right)$$

$$= - \sum_{I \in \mathcal{P}([n])} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m + \sum_{I \in \mathcal{P}([n+1]) \setminus \mathcal{P}([n])} (-1)^{(n+1)-|I|} \left(\sum_{i \in I} v_i \right)^m .$$

Adding $\sum_{I \in \mathcal{P}([n])} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m$ to both sides of this equality, we obtain

$$\sum_{I \in \mathcal{P}([n])} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m$$

$$= \sum_{I \in \mathcal{P}([n+1]) \setminus \mathcal{P}([n])} (-1)^{(n+1)-|I|} \left(\sum_{i \in I} v_i \right)^m$$

$$= \sum_{U \in \mathcal{P}([n])} (-1)^{(n+1)-|U \cup \{n+1\}|} \left(\sum_{i \in U \cup \{n+1\}} v_i \right)^m$$

$$\left(\begin{array}{c} \text{here, we have substituted } U \cup \{n+1\} \text{ for } I \text{ in the sum,} \\ \text{since the map } \mathcal{P}([n]) \to \mathcal{P}([n+1]) \setminus \mathcal{P}([n]) , \ U \mapsto U \cup \{n+1\} \\ \text{is a bijection} \end{array} \right)$$

$$= \sum_{I \in \mathcal{P}([n])} (-1)^{(n+1)-|I \cup \{n+1\}|} \left(\sum_{i \in I \cup \{n+1\}} v_i \right)^m \tag{1410}$$

(here, we have renamed the summation index U as I).

But each $I \in \mathcal{P}([n])$ satisfies

$$(-1)^{(n+1)-|I\cup\{n+1\}|}\left(\sum_{i\in I\cup\{n+1\}} v_i\right)^m = (-1)^{n-|I|}\left(w+\sum_{i\in I} v_i\right)^m \qquad (1411)$$

[627]. Thus, (1410) becomes

$$\begin{aligned}\sum_{I\in\mathcal{P}([n])}(-1)^{n-|I|}\left(\sum_{i\in I} v_i\right)^m &= \sum_{I\in\mathcal{P}([n])}\underbrace{(-1)^{(n+1)-|I\cup\{n+1\}|}\left(\sum_{i\in I\cup\{n+1\}} v_i\right)^m}_{\substack{=(-1)^{n-|I|}\left(w+\sum_{i\in I} v_i\right)^m \\ \text{(by (1411))}}} \\ &= \sum_{I\in\mathcal{P}([n])}(-1)^{n-|I|}\left(w+\sum_{i\in I} v_i\right)^m.\end{aligned}$$

In view of $\sum_{I\in\mathcal{P}([n])} = \sum_{I\subseteq[n]}$ (an equality between summation signs), this equality rewrites as

$$\sum_{I\subseteq[n]}(-1)^{n-|I|}\left(\sum_{i\in I} v_i\right)^m = \sum_{I\subseteq[n]}(-1)^{n-|I|}\left(w+\sum_{i\in I} v_i\right)^m.$$

This proves (1409).]

[627] *Proof of (1411):* Let $I \in \mathcal{P}([n])$. Thus, I is a subset of $[n]$ (since $\mathcal{P}([n])$ is the set of all subsets of $[n]$). In other words, $I \subseteq [n]$. Hence, $n+1 \notin I$. Therefore, $|I\cup\{n+1\}| = |I|+1$ and $I \setminus \{n+1\} = I$. Therefore,

$$(n+1) - \underbrace{|I\cup\{n+1\}|}_{=|I|+1} = (n+1) - (|I|+1) = n - |I|.$$

Furthermore, the sets I and $\{n+1\}$ are disjoint (since $n+1 \notin I$). Thus,

$$\sum_{i\in I\cup\{n+1\}} v_i = \sum_{i\in I} v_i + \underbrace{\sum_{i\in\{n+1\}} v_i}_{=v_{n+1}=w} = \sum_{i\in I} v_i + w = w + \sum_{i\in I} v_i.$$

Thus,

$$\underbrace{(-1)^{(n+1)-|I\cup\{n+1\}|}}_{\substack{=(-1)^{n-|I|} \\ \text{(since } (n+1)-|I\cup\{n+1\}|=n-|I|)}}\left(\underbrace{\sum_{i\in I\cup\{n+1\}} v_i}_{=w+\sum_{i\in I} v_i}\right)^m = (-1)^{n-|I|}\left(w+\sum_{i\in I} v_i\right)^m.$$

This proves (1411).

(a) Let $m \in \{0, 1, \ldots, n-1\}$ and $w \in \mathbb{L}$. We have $m \in \{0, 1, \ldots, n-1\} \subseteq \{0, 1, \ldots, n\}$. Hence, (1409) yields

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(w + \sum_{i \in I} v_i \right)^m = \sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^m = 0$$

(by Exercise 6.50 **(b)**). This solves Exercise 6.51 **(a)**.

(b) Let $w \in \mathbb{L}$. We have $n \in \{0, 1, \ldots, n\}$. Hence, (1409) (applied to $m = n$) yields

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(w + \sum_{i \in I} v_i \right)^n = \sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i \right)^n = \sum_{\sigma \in S_n} v_{\sigma(1)} v_{\sigma(2)} \cdots v_{\sigma(n)}$$

(by Exercise 6.50 **(c)**). This solves Exercise 6.51 **(b)**.

Next, let us prepare for the solutions of parts **(c)** and **(d)** of Exercise 6.51. Set $q = \sum_{i \in [n]} v_i$. Then, each subset I of $[n]$ satisfies

$$\begin{aligned} q = \sum_{i \in [n]} v_i &= \underbrace{\sum_{\substack{i \in [n]; \\ i \in I}}}_{\substack{= \sum_{i \in I} \\ (\text{since } I \subseteq [n])}} v_i + \underbrace{\sum_{\substack{i \in [n]; \\ i \notin I}}}_{= \sum_{i \in [n] \setminus I}} v_i \\ &\qquad (\text{since each } i \in [n] \text{ satisfies either } i \in I \text{ or } i \notin I \text{ (but not both)}) \\ &= \sum_{i \in I} v_i + \sum_{i \in [n] \setminus I} v_i \end{aligned}$$

and thus

$$\begin{aligned} - \underbrace{q}_{= \sum_{i \in I} v_i + \sum_{i \in [n] \setminus I} v_i} + \underbrace{\sum_{i \in I} 2 v_i}_{= 2 \sum_{i \in I} v_i} &= - \left(\sum_{i \in I} v_i + \sum_{i \in [n] \setminus I} v_i \right) + 2 \sum_{i \in I} v_i \\ &= \sum_{i \in I} v_i - \sum_{i \in [n] \setminus I} v_i. \qquad (1412) \end{aligned}$$

(c) Let $m \in \{0, 1, \ldots, n-1\}$. Then, Exercise 6.51 **(a)** (applied to $2v_i$ and $-q$ instead of v_i and w) yields

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(-q + \sum_{i \in I} 2 v_i \right)^m = 0.$$

Using (1412), we can rewrite this as

$$\sum_{I \subseteq [n]} (-1)^{n-|I|} \left(\sum_{i \in I} v_i - \sum_{i \in [n] \setminus I} v_i \right)^m = 0.$$

This solves Exercise 6.51 **(c)**.

(d) Exercise 6.51 **(b)** (applied to $2v_i$ and $-q$ instead of v_i and w) yields

$$\sum_{I\subseteq[n]}(-1)^{n-|I|}\left(-q+\sum_{i\in I}2v_i\right)^n=\sum_{\sigma\in S_n}\underbrace{\left(2v_{\sigma(1)}\right)\left(2v_{\sigma(2)}\right)\cdots\left(2v_{\sigma(n)}\right)}_{=2^n\cdot v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)}}$$
$$=2^n\sum_{\sigma\in S_n}v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)}.$$

Using (1412), we can rewrite this as

$$\sum_{I\subseteq[n]}(-1)^{n-|I|}\left(\sum_{i\in I}v_i-\sum_{i\in[n]\setminus I}v_i\right)^n=2^n\sum_{\sigma\in S_n}v_{\sigma(1)}v_{\sigma(2)}\cdots v_{\sigma(n)}.$$

This solves Exercise 6.51 **(d)**. □

7.120. Solution to Exercise 6.52

In this section, we shall use the same notations that we have introduced in Section 7.118.

We start by proving a simple restatement of Lemma 7.220:

Lemma 7.224. Let $\mathbb{L}$ be a noncommutative ring.

Let $n\in\mathbb{N}$. Let I be a finite set. For every $i\in I$ and every $j\in[n]$, we let $b_{i,j}$ be an element of $\mathbb{L}$. Then,

$$\prod_{j=1}^n\sum_{i\in I}b_{i,j}=\sum_{f:[n]\to I}\prod_{j=1}^n b_{f(j),j}.$$

Proof of Lemma 7.224. Define an $m\in\mathbb{N}$ by $m=|I|$. (This is well-defined, since I is a finite set.) Now, $|[m]|=m=|I|$. Hence, there exists a bijection $\varphi:[m]\to I$. Consider such a φ.

Lemma 7.220 (applied to $\mathbb{K}=\mathbb{L}$ and $p_{i,k}=b_{\varphi(k),i}$) yields

$$\prod_{i=1}^n\sum_{k=1}^m b_{\varphi(k),i}=\sum_{\kappa:[n]\to[m]}\prod_{i=1}^n b_{\varphi(\kappa(i)),i}. \tag{1413}$$

Every $i\in\{1,2,\ldots,n\}$ satisfies

$$\sum_{k=1}^m b_{\varphi(k),i}=\sum_{t\in I}b_{t,i} \tag{1414}$$

[628].

Hence,

$$\prod_{i=1}^{n}\underbrace{\sum_{k=1}^{m}b_{\varphi(k),i}}_{\substack{=\sum_{t\in I}b_{t,i}\\ \text{(by (1414))}}}=\prod_{i=1}^{n}\sum_{t\in I}b_{t,i}=\prod_{j=1}^{n}\underbrace{\sum_{t\in I}b_{t,j}}_{\substack{=\sum_{i\in I}b_{i,j}\\ \text{(here, we have renamed the}\\ \text{summation index } t \text{ as } i)}}$$

$$\left(\begin{array}{c}\text{here, we have renamed the index } i \text{ as } j\\ \text{in the product}\end{array}\right)$$

$$=\prod_{j=1}^{n}\sum_{i\in I}b_{i,j}. \tag{1415}$$

Comparing this with (1413), we find

$$\prod_{j=1}^{n}\sum_{i\in I}b_{i,j}=\sum_{\kappa:[n]\to[m]}\prod_{i=1}^{n}b_{\varphi(\kappa(i)),i}. \tag{1416}$$

The map φ is a bijection, and thus its inverse φ^{-1} exists. Hence, the map

$$[m]^{[n]}\to I^{[n]},\qquad \kappa\mapsto\varphi\circ\kappa$$

is a bijection (in fact, its inverse is the map $I^{[n]}\to[m]^{[n]}$, $f\mapsto\varphi^{-1}\circ f$).

Therefore, we can substitute $\varphi\circ\kappa$ for f in the sum $\sum_{f\in I^{[n]}}\prod_{j=1}^{n}b_{f(j),j}$. We thus obtain

$$\sum_{f\in I^{[n]}}\prod_{j=1}^{n}b_{f(j),j}=\underbrace{\sum_{\kappa\in[m]^{[n]}}}_{=\sum_{\kappa:[n]\to[m]}}\prod_{j=1}^{n}\underbrace{b_{(\varphi\circ\kappa)(j),j}}_{\substack{=b_{\varphi(\kappa(j)),j}\\ \text{(since } (\varphi\circ\kappa)(j)=\varphi(\kappa(j)))}}=\sum_{\kappa:[n]\to[m]}\underbrace{\prod_{j=1}^{n}b_{\varphi(\kappa(j)),j}}_{\substack{=\prod_{i=1}^{n}b_{\varphi(\kappa(i)),i}\\ \text{(here, we have renamed}\\ \text{the index } j \text{ as } i \text{ in}\\ \text{the product)}}}$$

$$=\sum_{\kappa:[n]\to[m]}\prod_{i=1}^{n}b_{\varphi(\kappa(i)),i}.$$

[628] *Proof of (1414):* Let $i\in\{1,2,\ldots,n\}$. Recall that $\varphi:[m]\to I$ is a bijection. Hence, we can substitute $\varphi(k)$ for t in the sum $\sum_{t\in I}b_{t,i}$. We thus obtain

$$\sum_{t\in I}b_{t,i}=\underbrace{\sum_{k\in[m]}}_{\substack{=\sum_{k\in\{1,2,\ldots,m\}}\\ \text{(since } [m]=\{1,2,\ldots,m\})}}b_{\varphi(k),i}=\underbrace{\sum_{k\in\{1,2,\ldots,m\}}}_{=\sum_{k=1}^{m}}b_{\varphi(k),i}=\sum_{k=1}^{m}b_{\varphi(k),i}.$$

This proves (1414).

Comparing this with (1416), we obtain

$$\prod_{j=1}^{n}\sum_{i\in I}b_{i,j}=\underbrace{\sum_{f\in I^{[n]}}}_{=\sum_{f:[n]\to I}}\prod_{j=1}^{n}b_{f(j),j}=\sum_{f:[n]\to I}\prod_{j=1}^{n}b_{f(j),j}.$$

This proves Lemma 7.224. □

Next, we state an analogue of Lemma 7.222:

Lemma 7.225. Let $\mathbb{L}$ be a noncommutative ring. Let G be a finite set. Let $n\in\mathbb{N}$. For each $i\in G$ and $j\in[n]$, let $b_{i,j}$ be an element of $\mathbb{L}$. Let I be a subset of G. Then,

$$\prod_{j=1}^{n}\sum_{i\in I}b_{i,j}=\sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}}\prod_{j=1}^{n}b_{f(j),j}.$$

Proof of Lemma 7.225. For each $j\in[n]$, we have

$$\begin{aligned}
&\sum_{i\in G}[i\in I]\,b_{i,j}\\
&=\underbrace{\sum_{\substack{i\in G;\\ i\in I}}}_{\substack{=\sum_{i\in I}\\ \text{(since } I\subseteq G)}}\underbrace{[i\in I]}_{\substack{=1\\ \text{(since } i\in I \text{ is true)}}}b_{i,j}+\sum_{\substack{i\in G;\\ i\notin I}}\underbrace{[i\in I]}_{\substack{=0\\ \text{(since } i\in I \text{ is false}\\ \text{(because } i\notin I))}}b_{i,j}\\
&\qquad\text{(since each } i\in G \text{ satisfies either } i\in I \text{ or } i\notin I \text{ (but not both))}\\
&=\sum_{i\in I}1b_{i,j}+\underbrace{\sum_{\substack{i\in G;\\ i\notin I}}0b_{i,j}}_{=0}=\sum_{i\in I}1b_{i,j}=\sum_{i\in I}b_{i,j}.
\end{aligned}\tag{1417}$$

Hence,

$$\prod_{j=1}^{n}\underbrace{\sum_{i\in G}[i\in I]\,b_{i,j}}_{\substack{=\sum_{i\in I}b_{i,j}\\ \text{(by (1417))}}}=\prod_{j=1}^{n}\sum_{i\in I}b_{i,j}.$$

Thus,

$$\prod_{j=1}^{n}\sum_{i\in I}b_{i,j}=\prod_{j=1}^{n}\sum_{i\in G}[i\in I]\,b_{i,j}=\sum_{f:[n]\to G}\prod_{j=1}^{n}\left([f(j)\in I]\,b_{f(j),j}\right)\tag{1418}$$

(by Lemma 7.224 (applied to G and $[i\in I]\,b_{i,j}$ instead of I and $b_{i,j}$)).

Now, fix any map $f : [n] \to G$. Then,

$$\prod_{j=1}^{n} \left([f(j) \in I] \, b_{f(j),j} \right) = \left([f(1) \in I] \, b_{f(1),1} \right) \left([f(2) \in I] \, b_{f(2),2} \right) \cdots \left([f(n) \in I] \, b_{f(n),n} \right). \tag{1419}$$

The factors $[f(1) \in I], [f(2) \in I], \ldots, [f(n) \in I]$ on the right hand side of this equality are integers, and therefore can be freely moved within the product (even though $\mathbb{L}$ is not necessarily commutative). In particular, we can move them to the front of the product. Thus, we find

$$\begin{aligned} & \left([f(1) \in I] \, b_{f(1),1} \right) \left([f(2) \in I] \, b_{f(2),2} \right) \cdots \left([f(n) \in I] \, b_{f(n),n} \right) \\ &= \underbrace{([f(1) \in I] [f(2) \in I] \cdots [f(n) \in I])}_{= \prod_{i=1}^{n} [f(i) \in I]} \left(b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n} \right) \\ &= \left(\prod_{i=1}^{n} [f(i) \in I] \right) \left(b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n} \right). \end{aligned}$$

Hence, (1419) rewrites as

$$\prod_{j=1}^{n} \left([f(j) \in I] \, b_{f(j),j} \right) = \left(\prod_{i=1}^{n} [f(i) \in I] \right) \left(b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n} \right). \tag{1420}$$

But Lemma 7.221 (applied to n instead of m) yields

$$\prod_{i=1}^{n} [f(i) \in I] = [f([n]) \subseteq I]. \tag{1421}$$

Hence, (1420) becomes

$$\begin{aligned} \prod_{j=1}^{n} \left([f(j) \in I] \, b_{f(j),j} \right) &= \underbrace{\left(\prod_{i=1}^{n} [f(i) \in I] \right)}_{\substack{= [f([n]) \subseteq I] \\ \text{(by (1421))}}} \left(b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n} \right) \\ &= [f([n]) \subseteq I] \left(b_{f(1),1} b_{f(2),2} \cdots b_{f(n),n} \right). \end{aligned} \tag{1422}$$

Now, forget that we fixed f. We thus have proven (1422) for each map $f : [n] \to G$.

Now, (1418) becomes

$$\prod_{j=1}^{n}\sum_{i\in I} b_{i,j} = \sum_{f:[n]\to G} \underbrace{\prod_{j=1}^{n}\left([f(j)\in I]\, b_{f(j),j}\right)}_{\substack{=[f([n])\subseteq I]\left(b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}\right) \\ \text{(by (1422))}}}$$

$$= \sum_{f:[n]\to G} [f([n])\subseteq I]\left(b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}\right)$$

$$= \sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}} \underbrace{[f([n])\subseteq I]}_{\substack{=1\\ \text{(since } f([n])\subseteq I)}} \left(b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}\right) + \sum_{\substack{f:[n]\to G;\\ \text{not } f([n])\subseteq I}} \underbrace{[f([n])\subseteq I]}_{\substack{=0\\ \text{(since we don't have } f([n])\subseteq I)}} \left(b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}\right)$$

$$\left(\begin{array}{c}\text{since each map } f:[n]\to G \text{ satisfies either } f([n])\subseteq I\\ \text{or } (\text{not } f([n])\subseteq I) \text{ (but not both)}\end{array}\right)$$

$$= \sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n} + \underbrace{\sum_{\substack{f:[n]\to G;\\ \text{not } f([n])\subseteq I}} 0\left(b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}\right)}_{=0}$$

$$= \sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}} \underbrace{b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}}_{=\prod_{j=1}^{n} b_{f(j),j}} = \sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}} \prod_{j=1}^{n} b_{f(j),j}.$$

This proves Lemma 7.225. □

The next result we shall use can be viewed as a slight generalization of Lemma 7.223:

Lemma 7.226. Let G be a finite set. Let H and T be two subsets of G. Then,

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ T\subseteq I}} (-1)^{|I|} = (-1)^{|G|}\,[G\setminus H\subseteq T].$$

Proof of Lemma 7.226. If I is a subset of G, then the statement $(H\subseteq I \text{ and } T\subseteq I)$ is equivalent to the statement $(H\cup T\subseteq I)$ [629]. Hence, we have the following equality of summation signs:

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ T\subseteq I}} = \sum_{\substack{I\subseteq G;\\ H\cup T\subseteq I}}. \qquad (1423)$$

[629] In fact, this equivalence is a basic fact of set theory, which holds for any set I.

Let $R = H \cup T$. Then, R is a subset of G (since H and T are subsets of G). Moreover, the equality (1423) becomes

$$\sum_{\substack{I \subseteq G; \\ H \subseteq I; \\ T \subseteq I}} = \sum_{\substack{I \subseteq G; \\ H \cup T \subseteq I}} = \sum_{\substack{I \subseteq G; \\ R \subseteq I}} \tag{1424}$$

(since $H \cup T = R$).

We are in one of the following two cases:

Case 1: We have $G \setminus H \not\subseteq T$.

Case 2: We have $G \setminus H \subseteq T$.

Let us first consider Case 1. In this case, we have $G \setminus H \not\subseteq T$. Hence, there exists some $g \in G \setminus H$ such that $g \notin T$. Consider this g.

We have $g \in G \setminus H$. In other words, $g \in G$ and $g \notin H$. We have $g \notin H \cup T$ (since $g \notin H$ and $g \notin T$). In other words, $g \notin R$ (since $R = H \cup T$). Also, $\{g\}$ is a subset of G (since $g \in G$).

Now, the map

$$\{I \subseteq G \mid R \subseteq I \text{ and } g \in I\} \to \{I \subseteq G \mid R \subseteq I \text{ and } g \notin I\}, \\ J \mapsto J \setminus \{g\} \tag{1425}$$

is well-defined[630] and bijective[631].

[630] This is because every $J \in \{I \subseteq G \mid R \subseteq I \text{ and } g \in I\}$ satisfies $J \setminus \{g\} \in \{I \subseteq G \mid R \subseteq I \text{ and } g \notin I\}$. (Proving this is straightforward using the facts that $g \in G$ and $g \notin R$.)

[631] Indeed, its inverse is the map

$$\{I \subseteq G \mid R \subseteq I \text{ and } g \notin I\} \to \{I \subseteq G \mid R \subseteq I \text{ and } g \in I\}, \\ J \mapsto J \cup \{g\}$$

(which, again, is well-defined because $g \in G$).

Hence,

$$\underbrace{\sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\notin J}}}_{=\sum_{J\in\{I\subseteq G\ \mid\ R\subseteq I\text{ and }g\notin I\}}} (-1)^{|J|}$$

$$= \sum_{J\in\{I\subseteq G\ \mid\ R\subseteq I\text{ and }g\notin I\}} (-1)^{|J|} = \underbrace{\sum_{J\in\{I\subseteq G\ \mid\ R\subseteq I\text{ and }g\in I\}}}_{=\sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}}} (-1)^{|J\setminus\{g\}|}$$

$$\left(\begin{array}{c}\text{here, we have substituted } J\setminus\{g\} \text{ for } J \text{ in the sum,}\\ \text{since the map (1425) is bijective}\end{array}\right)$$

$$= \sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} \underbrace{(-1)^{|J\setminus\{g\}|}}_{\substack{=(-1)^{|J|-1}\\ (\text{since } |J\setminus\{g\}|=|J|-1\\ (\text{because } g\in J))}} = \sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} \underbrace{(-1)^{|J|-1}}_{=-(-1)^{|J|}} = \sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} \left(-(-1)^{|J|}\right)$$

$$= -\sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} (-1)^{|J|}. \qquad (1426)$$

But (1424) yields

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ T\subseteq I}} (-1)^{|I|} = \sum_{\substack{I\subseteq G;\\ R\subseteq I}} (-1)^{|I|} = \sum_{\substack{J\subseteq G;\\ R\subseteq J}} (-1)^{|J|} \qquad \left(\begin{array}{c}\text{here, we have renamed the}\\ \text{summation index } I \text{ as } J\end{array}\right)$$

$$= \sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} (-1)^{|J|} + \underbrace{\sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\notin J}} (-1)^{|J|}}_{\substack{=-\sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} (-1)^{|J|}\\ (\text{by (1426)})}} = \sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} (-1)^{|J|} - \sum_{\substack{J\subseteq G;\\ R\subseteq J;\\ g\in J}} (-1)^{|J|} = 0.$$

Comparing this with $(-1)^{|G|} \underbrace{[G\setminus H\subseteq T]}_{\substack{=0\\ (\text{since } G\setminus H\not\subseteq T)}} = 0$, we obtain

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ T\subseteq I}} (-1)^{|I|} = (-1)^{|G|} [G\setminus H\subseteq T].$$

Hence, Lemma 7.226 is proven in Case 1.

Let us now consider Case 2. In this case, we have $G \setminus H \subseteq T$. Hence, $T \supseteq G \setminus H$. Now,

$$R = H \cup \underbrace{T}_{\supseteq G \setminus H} \supseteq H \cup (G \setminus H) = G$$

(since H is a subset of G). Combining this with $R \subseteq G$, we obtain $R = G$.

Now, the equality (1424) becomes

$$\begin{aligned} \sum_{\substack{I \subseteq G; \\ H \subseteq I; \\ T \subseteq I}} = \sum_{\substack{I \subseteq G; \\ R \subseteq I}} &= \sum_{\substack{I \subseteq G; \\ G \subseteq I}} \qquad (\text{since } R = G) \\ &= \sum_{\substack{I \subseteq G; \\ I = G}} \qquad \left(\begin{array}{c} \text{because for a subset } I \text{ of } G \text{, the} \\ \text{statement } (G \subseteq I) \text{ is equivalent to } (I = G) \end{array} \right). \end{aligned}$$

Hence,

$$\underbrace{\sum_{\substack{I \subseteq G; \\ H \subseteq I; \\ T \subseteq I}}}_{= \sum_{\substack{I \subseteq G; \\ I = G}}} (-1)^{|I|} = \sum_{\substack{I \subseteq G; \\ I = G}} (-1)^{|I|} = (-1)^{|G|} \qquad (\text{since } G \text{ is a subset of } G).$$

Comparing this with $(-1)^{|G|} \underbrace{[G \setminus H \subseteq T]}_{\substack{=1 \\ (\text{since } G \setminus H \subseteq T)}} = (-1)^{|G|}$, we obtain

$$\sum_{\substack{I \subseteq G; \\ H \subseteq I; \\ T \subseteq I}} (-1)^{|I|} = (-1)^{|G|} [G \setminus H \subseteq T].$$

Hence, Lemma 7.226 is proven in Case 2.

We have thus proven Lemma 7.226 in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Lemma 7.226 always holds. □

Solution to Exercise 6.52. **(a)** We have

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} \underbrace{b_{I,1}b_{I,2}\cdots b_{I,n}}_{=\prod_{j=1}^{n} b_{I,j}}$$

$$= \sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} \prod_{j=1}^{n} \underbrace{b_{I,j}}_{\substack{=\sum_{i\in I} b_{i,j} \\ \text{(by the definition of } b_{I,j})}} = \sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} \underbrace{\prod_{j=1}^{n}\sum_{i\in I} b_{i,j}}_{\substack{=\sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}} \prod_{j=1}^{n} b_{f(j),j} \\ \text{(by Lemma 7.225)}}}$$

$$= \sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} \sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}} \prod_{j=1}^{n} b_{f(j),j} = \underbrace{\sum_{\substack{I\subseteq G;\\ H\subseteq I}} \sum_{\substack{f:[n]\to G;\\ f([n])\subseteq I}}}_{=\sum_{f:[n]\to G} \sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ f([n])\subseteq I}}} (-1)^{|I|} \prod_{j=1}^{n} b_{f(j),j}$$

$$= \sum_{f:[n]\to G} \sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ f([n])\subseteq I}} (-1)^{|I|} \prod_{j=1}^{n} b_{f(j),j} = \sum_{f:[n]\to G} \underbrace{\left(\sum_{\substack{I\subseteq G;\\ H\subseteq I;\\ f([n])\subseteq I}} (-1)^{|I|} \right)}_{\substack{=(-1)^{|G|}[G\setminus H\subseteq f([n])] \\ \text{(by Lemma 7.226} \\ \text{(applied to } T=f([n])))}} \prod_{j=1}^{n} b_{f(j),j}$$

$$= \sum_{f:[n]\to G} (-1)^{|G|} [G\setminus H \subseteq f([n])] \prod_{j=1}^{n} b_{f(j),j}$$

$$= \sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} (-1)^{|G|} \underbrace{[G\setminus H \subseteq f([n])]}_{\substack{=1 \\ \text{(since } G\setminus H\subseteq f([n]))}} \prod_{j=1}^{n} b_{f(j),j}$$

$$+ \sum_{\substack{f:[n]\to G;\\ \text{not } G\setminus H\subseteq f([n])}} (-1)^{|G|} \underbrace{[G\setminus H \subseteq f([n])]}_{\substack{=0 \\ \text{(since we don't have } G\setminus H\subseteq f([n]))}} \prod_{j=1}^{n} b_{f(j),j}$$

$$\left(\begin{array}{c} \text{since each map } f:[n]\to G \text{ satisfies either } (G\setminus H\subseteq f([n])) \\ \text{or } (\text{not } G\setminus H\subseteq f([n])) \text{ (but not both)} \end{array} \right)$$

$$= \sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} (-1)^{|G|} \prod_{j=1}^{n} b_{f(j),j} + \underbrace{\sum_{\substack{f:[n]\to G;\\ \text{not } G\setminus H\subseteq f([n])}} (-1)^{|G|} 0 \prod_{j=1}^{n} b_{f(j),j}}_{=0}$$
$$= \sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} (-1)^{|G|} \prod_{j=1}^{n} b_{f(j),j} = (-1)^{|G|} \sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} \underbrace{\prod_{j=1}^{n} b_{f(j),j}}_{=b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}}$$
$$= (-1)^{|G|} \sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}.$$

This solves Exercise 6.52 **(a)**.

(b) Assume that $n < |G\setminus H|$. Then, there exists no $f : [n] \to G$ satisfying $G\setminus H \subseteq f([n])$ [632]. Hence, the sum

$$\sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}$$

is empty, and thus equals 0. In other words,

$$\sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n} = 0.$$

Now, Exercise 6.52 **(a)** yields

$$\sum_{\substack{I\subseteq G;\\ H\subseteq I}} (-1)^{|I|} b_{I,1}b_{I,2}\cdots b_{I,n} = (-1)^{|G|} \underbrace{\sum_{\substack{f:[n]\to G;\\ G\setminus H\subseteq f([n])}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}}_{=0} = 0.$$

This solves Exercise 6.52 **(b)**.

(c) Assume that $n = |G|$. For any map $f : [n] \to G$, we have the following logical equivalence:

$$(G\setminus\varnothing \subseteq f([n])) \iff (f \text{ is bijective})$$

[633]. Thus, we have the following equality of summation signs:

$$\sum_{\substack{f:[n]\to G;\\ G\setminus\varnothing\subseteq f([n])}} = \sum_{\substack{f:[n]\to G;\\ f \text{ is bijective}}} . \qquad (1427)$$

[632] *Proof.* Let $f : [n] \to G$ be such that $G\setminus H \subseteq f([n])$. We shall derive a contradiction.

Clearly, $[n]$ is a subset of the finite set $[n]$. Thus, Lemma 1.3 **(a)** (applied to $[n]$, G and $[n]$ instead of U, V and S) yields $|f([n])| \leq |[n]| = n$. But $G\setminus H \subseteq f([n])$; hence, $|G\setminus H| \leq |f([n])| \leq n$. Now, $n < |G\setminus H| \leq n$. This is absurd. Hence, we have obtained a contradiction.

Now, forget that we fixed f. We thus have found a contradiction for each $f : [n] \to G$ satisfying $G\setminus H \subseteq f([n])$. Hence, there exists no $f : [n] \to G$ satisfying $G\setminus H \subseteq f([n])$.

[633] *Proof.* We know that $[n]$ and G are two finite sets such that $|[n]| \leq |G|$ (since $|[n]| = n \leq n = |G|$). Thus, Lemma 1.4 (applied to $U = [n]$ and $V = G$) shows that we have the following logical

Also, each subset I of G satisfies $\varnothing \subseteq I$. Hence, we have the following equality of summation signs:

$$\sum_{\substack{I\subseteq G;\\ \varnothing\subseteq I}} = \sum_{I\subseteq G} . \tag{1428}$$

Now, $\varnothing$ is a subset of G. Hence, Exercise 6.52 **(a)** (applied to $\varnothing$ instead of H) yields

$$\begin{aligned}\sum_{\substack{I\subseteq G;\\ \varnothing\subseteq I}} (-1)^{|I|} b_{I,1}b_{I,2}\cdots b_{I,n} &= (-1)^{|G|} \underbrace{\sum_{\substack{f:[n]\to G;\\ G\setminus\varnothing\subseteq f([n])}}}_{\substack{=\sum_{\substack{f:[n]\to G;\\ f\text{ is bijective}}}\\ \text{(by (1427))}}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n} \\ &= (-1)^{|G|} \sum_{\substack{f:[n]\to G;\\ f\text{ is bijective}}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}. \end{aligned}\tag{1429}$$

equivalence:

$$(f\text{ is surjective}) \iff (f\text{ is bijective}) .$$

But we have the following logical equivalence:

$$\begin{aligned}\left(\underbrace{G\setminus\varnothing}_{=G}\subseteq f([n])\right) &\iff (G\subseteq f([n])) \iff (G = f([n])) \\ &\qquad \left(\text{since } f([n]) \text{ is a subset of } G\right) \\ &\iff (f\text{ is surjective}) \iff (f\text{ is bijective}) .\end{aligned}$$

Qed.

Now,

$$\sum_{I\subseteq G}\underbrace{(-1)^{|G\setminus I|}}_{\substack{=(-1)^{|G|-|I|}\\ \text{(since } |G\setminus I|=|G|-|I|\\ \text{(because } I\subseteq G))}} b_{I,1}b_{I,2}\cdots b_{I,n}$$
$$=\sum_{I\subseteq G}\underbrace{(-1)^{|G|-|I|}}_{\substack{=(-1)^{|G|+|I|}\\ \text{(since } |G|-|I|\equiv|G|+|I| \operatorname{mod} 2)}} b_{I,1}b_{I,2}\cdots b_{I,n}=\underbrace{\sum_{I\subseteq G}}_{\substack{=\sum_{\substack{I\subseteq G;\\ \varnothing\subseteq I}}\\ \text{(by (1428))}}}\underbrace{(-1)^{|G|+|I|}}_{=(-1)^{|G|}(-1)^{|I|}} b_{I,1}b_{I,2}\cdots b_{I,n}$$
$$=\sum_{\substack{I\subseteq G;\\ \varnothing\subseteq I}}(-1)^{|G|}(-1)^{|I|}b_{I,1}b_{I,2}\cdots b_{I,n}=(-1)^{|G|}\underbrace{\sum_{\substack{I\subseteq G;\\ \varnothing\subseteq I}}(-1)^{|I|}b_{I,1}b_{I,2}\cdots b_{I,n}}_{\substack{=(-1)^{|G|}\sum_{\substack{f:[n]\to G;\\ f\text{ is bijective}}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}\\ \text{(by (1429))}}}$$
$$=\underbrace{(-1)^{|G|}(-1)^{|G|}}_{\substack{=((-1)(-1))^{|G|}=1^{|G|}\\ \text{(since } (-1)(-1)=1)}}\sum_{\substack{f:[n]\to G;\\ f\text{ is bijective}}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}$$
$$=\underbrace{1^{|G|}}_{=1}\sum_{\substack{f:[n]\to G;\\ f\text{ is bijective}}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}$$
$$=\sum_{\substack{f:[n]\to G;\\ f\text{ is bijective}}} b_{f(1),1}b_{f(2),2}\cdots b_{f(n),n}.$$

This solves Exercise 6.52 **(c)**. □

7.121. Solution to Exercise 6.53

In this section, we shall use the same notations that we have introduced in Section 7.118.

Let us begin by stating a simple corollary from Exercise 6.52 **(b)**:

Lemma 7.227. Let G be a finite set. Let $n\in\mathbb{N}$ be such that $n<|G|$.
For each $i\in G$ and $j\in[n]$, let $b_{i,j}$ be an element of $\mathbb{K}$. Then,

$$\sum_{I\subseteq G}(-1)^{|I|}\prod_{i=1}^{n}\sum_{g\in I}b_{g,i}=0.$$

Proof of Lemma 7.227. Clearly, $\varnothing$ is a subset of G. Also, $G\setminus\varnothing=G$, so that $|G\setminus\varnothing|=|G|$. Thus, $n<|G|=|G\setminus\varnothing|$.

For each $j \in [n]$ and each subset I of G, we define an element $b_{I,j} \in \mathbb{K}$ by $b_{I,j} = \sum_{i \in I} b_{i,j}$. Thus, for each $j \in [n]$ and each subset I of G, we have

$$b_{I,j} = \sum_{i \in I} b_{i,j} = \sum_{g \in I} b_{g,j} \tag{1430}$$

(here, we have substituted g for i in the sum).

Now, Exercise 6.52 **(b)** (applied to $\mathbb{K}$ and $\varnothing$ instead of $\mathbb{L}$ and H) yields

$$\sum_{\substack{I \subseteq G; \\ \varnothing \subseteq I}} (-1)^{|I|} b_{I,1} b_{I,2} \cdots b_{I,n} = 0. \tag{1431}$$

But each subset I of G satisfies $\varnothing \subseteq I$. Hence, we have the following equality of summation signs:

$$\sum_{\substack{I \subseteq G; \\ \varnothing \subseteq I}} = \sum_{I \subseteq G} .$$

Thus, the equality (1431) rewrites as follows:

$$\sum_{I \subseteq G} (-1)^{|I|} b_{I,1} b_{I,2} \cdots b_{I,n} = 0.$$

Hence,

$$0 = \sum_{I \subseteq G} (-1)^{|I|} \underbrace{b_{I,1} b_{I,2} \cdots b_{I,n}}_{= \prod_{i=1}^{n} b_{I,i}} = \sum_{I \subseteq G} (-1)^{|I|} \prod_{i=1}^{n} \underbrace{b_{I,i}}_{\substack{= \sum_{g \in I} b_{g,i} \\ \text{(by (1430) (applied to } j=i\text{))}}}$$
$$= \sum_{I \subseteq G} (-1)^{|I|} \prod_{i=1}^{n} \sum_{g \in I} b_{g,i}.$$

This proves Lemma 7.227. $\square$

Solution to Exercise 6.53. For each $g \in G$, let us write the $n \times n$-matrix A_g in the form

$$A_g = \left(a_{g,i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} . \tag{1432}$$

Let I be any subset of G. Then,

$$\sum_{g \in I} \underbrace{A_g}_{\substack{= \left(a_{g,i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \\ \text{(by (1432))}}} = \sum_{g \in I} \left(a_{g,i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(\sum_{g \in I} a_{g,i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} .$$

Hence,

$$\det\left(\sum_{g \in I} A_g\right) = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} \sum_{g \in I} a_{g,i,\sigma(i)}$$

(by (341) (applied to $\sum_{g\in I} A_g$ and $\sum_{g\in I} a_{g,i,j}$ instead of A and $a_{i,j}$)). In view of

$$\sum_{g\in I} A_g = \sum_{i\in I} A_i \qquad \text{(here, we have renamed the summation index } g \text{ as } i\text{)},$$

this rewrites as follows:

$$\det\left(\sum_{i\in I} A_i\right) = \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n \sum_{g\in I} a_{g,i,\sigma(i)}. \tag{1433}$$

Now, forget that we fixed I. We thus have proven (1433) for each subset I of G. Now,

$$\begin{aligned}
&\sum_{I\subseteq G} (-1)^{|I|} \underbrace{\det\left(\sum_{i\in I} A_i\right)}_{\substack{=\sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n \sum_{g\in I} a_{g,i,\sigma(i)} \\ \text{(by (1433))}}} \\
&= \sum_{I\subseteq G} (-1)^{|I|} \sum_{\sigma\in S_n} (-1)^\sigma \prod_{i=1}^n \sum_{g\in I} a_{g,i,\sigma(i)} = \underbrace{\sum_{I\subseteq G} \sum_{\sigma\in S_n}}_{=\sum_{\sigma\in S_n} \sum_{I\subseteq G}} (-1)^{|I|} (-1)^\sigma \prod_{i=1}^n \sum_{g\in I} a_{g,i,\sigma(i)} \\
&= \sum_{\sigma\in S_n} \sum_{I\subseteq G} (-1)^{|I|} (-1)^\sigma \prod_{i=1}^n \sum_{g\in I} a_{g,i,\sigma(i)} = \sum_{\sigma\in S_n} (-1)^\sigma \underbrace{\sum_{I\subseteq G} (-1)^{|I|} \prod_{i=1}^n \sum_{g\in I} a_{g,i,\sigma(i)}}_{\substack{=0 \\ \text{(by Lemma 7.227} \\ \text{(applied to } b_{i,j}=a_{i,j,\sigma(j)}\text{))}}} \\
&= \sum_{\sigma\in S_n} (-1)^\sigma 0 = 0.
\end{aligned}$$

This solves Exercise 6.53. □

7.122. Solution to Exercise 6.54

We shall first prove some identities in preparation for the solution to Exercise 6.54.

First, we introduce a notation: For every $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$. We shall also use the notation introduced in Definition 6.81 throughout this section.

Let us now state a simple corollary of Lemma 6.22:

Corollary 7.228. Let $n \in \mathbb{N}$. Let $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq n}$ be an $n\times n$-matrix. Let $\sigma \in S_n$. Then,

$$\left(\sum_{i=1}^n a_{i,\sigma(i)}\right)^k = \sum_{\kappa:[k]\to[n]} \prod_{i=1}^k a_{\kappa(i),\sigma(\kappa(i))}$$

for every $k \in \mathbb{N}$.

Proof of Corollary 7.228. Let $g \in \mathbb{N}$. Then, Lemma 6.22 (applied to g, n and $a_{k,\sigma(k)}$ instead of n, m and $p_{i,k}$) yields

$$\prod_{i=1}^{g} \sum_{k=1}^{n} a_{k,\sigma(k)} = \sum_{\kappa:[g]\to[n]} \prod_{i=1}^{g} a_{\kappa(i),\sigma(\kappa(i))}.$$

Thus,

$$\sum_{\kappa:[g]\to[n]} \prod_{i=1}^{g} a_{\kappa(i),\sigma(\kappa(i))} = \prod_{i=1}^{g} \sum_{k=1}^{n} a_{k,\sigma(k)} = \left(\sum_{k=1}^{n} a_{k,\sigma(k)}\right)^{g} = \left(\sum_{i=1}^{n} a_{i,\sigma(i)}\right)^{g} \qquad (1434)$$

(here, we have renamed the summation index k as i in the sum).

Now, forget that we fixed g. We thus have proven the identity (1434) for each $g \in \mathbb{N}$.

Now, fix $k \in \mathbb{N}$. Then, (1434) (applied to $g = k$) yields $\sum_{\kappa:[k]\to[n]} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))} = \left(\sum_{i=1}^{n} a_{i,\sigma(i)}\right)^{k}$. This proves Corollary 7.228. □

Next, we state a simple lemma:

Lemma 7.229. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Let $\kappa : [k] \to [n]$ be a map such that $|\kappa([k])| < n - 1$. Then,

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))} = 0.$$

Proof of Lemma 7.229. We have $|[n] \setminus \kappa([k])| \geq 2$ [634]. Hence, the set $[n] \setminus \kappa([k])$ has at least 2 elements. In other words, the set $[n] \setminus \kappa([k])$ contains two distinct elements. Choose two such elements, and denote them by a and b. Thus, a and b are two distinct elements of the set $[n] \setminus \kappa([k])$.

The elements a and b both belong to $[n]$ (since $a \in [n] \setminus \kappa([k]) \subseteq [n]$ and $b \in [n] \setminus \kappa([k]) \subseteq [n]$).

Thus, a and b are two distinct elements of $[n] = \{1, 2, \ldots, n\}$. Hence, a transposition $t_{a,b} \in S_n$ is defined (according to Definition 5.29). This transposition satisfies $t_{a,b} \circ \kappa = \kappa$ [635].

[634] *Proof.* We have $|\kappa([k])| < n - 1$. Thus, $|\kappa([k])| \leq (n-1) - 1$ (since both $|\kappa([k])|$ and $n - 1$ are integers). Hence, $|\kappa([k])| \leq (n-1) - 1 = n - 2$.

But $\kappa([k]) \subseteq [n]$. Hence, $|[n] \setminus \kappa([k])| = \underbrace{|[n]|}_{=n} - \underbrace{|\kappa([k])|}_{\leq n-2} \geq n - (n-2) = 2$. Qed.

[635] *Proof.* We are going to show that every $i \in [k]$ satisfies $(t_{a,b} \circ \kappa)(i) = \kappa(i)$.

Let A_n be the set of all even permutations in S_n. Let C_n be the set of all odd permutations in S_n.

We have $\sigma \circ t_{a,b} \in C_n$ for every $\sigma \in A_n$ [636]. Hence, we can define a map $\Phi : A_n \to C_n$ by

$$\Phi(\sigma) = \sigma \circ t_{a,b} \qquad \text{for every } \sigma \in A_n.$$

Consider this map Φ. Furthermore, we have $\sigma \circ (t_{a,b})^{-1} \in A_n$ for every $\sigma \in C_n$ [637]. Thus, we can define a map $\Psi : C_n \to A_n$ by

$$\Psi(\sigma) = \sigma \circ (t_{a,b})^{-1} \qquad \text{for every } \sigma \in C_n.$$

Consider this map Ψ.

The maps Φ and Ψ are mutually inverse[638]. Hence, the map Φ is a bijection. Moreover, every $\sigma \in A_n$ and $i \in \{1, 2, \ldots, k\}$ satisfy

$$a_{\kappa(i),(\Phi(\sigma))(\kappa(i))} = a_{\kappa(i),\sigma(\kappa(i))} \tag{1435}$$

[639].

So let $i \in [k]$. We shall show that $(t_{a,b} \circ \kappa)(i) = \kappa(i)$.

We have $t_{a,b}(j) = j$ for every $j \in [n] \setminus \{a, b\}$ (by the definition of $t_{a,b}$).

But a and b are two elements of the set $[n] \setminus \kappa([k])$. Hence, $\{a, b\} \subseteq [n] \setminus \kappa([k])$. If we had $\kappa(i) \in \{a, b\}$, then we would thus have $\kappa(i) \in \{a, b\} \subseteq [n] \setminus \kappa([k])$, which would contradict $\kappa(i) \in \kappa([k])$. Hence, we cannot have $\kappa(i) \in \{a, b\}$. Thus, we have $\kappa(i) \notin \{a, b\}$.

Combining $\kappa(i) \in [n]$ with $\kappa(i) \notin \{a, b\}$, we obtain $\kappa(i) \in [n] \setminus \{a, b\}$. But recall that $t_{a,b}(j) = j$ for every $j \in [n] \setminus \{a, b\}$. Applying this to $j = \kappa(i)$, we obtain $t_{a,b}(\kappa(i)) = \kappa(i)$ (since $\kappa(i) \in [n] \setminus \{a, b\}$). Hence, $(t_{a,b} \circ \kappa)(i) = t_{a,b}(\kappa(i)) = \kappa(i)$.

Now, let us forget that we fixed i. We thus have shown that $(t_{a,b} \circ \kappa)(i) = \kappa(i)$ for every $i \in [k]$. In other words, $t_{a,b} \circ \kappa = \kappa$, qed.

[636]We have already proven this during our proof of Lemma 6.17 **(b)**.

[637]We have already proven this during our proof of Lemma 6.17 **(b)**.

[638]We have already proven this during our proof of Lemma 6.17 **(b)**.

[639]*Proof of (1435):* Let $\sigma \in A_n$ and $i \in \{1, 2, \ldots, k\}$. Thus, $i \in \{1, 2, \ldots, k\} = [k]$. Now,

$$\Bigg(\underbrace{\Phi(\sigma)}_{=\sigma \circ t_{a,b}}\Bigg)(\kappa(i)) = (\sigma \circ t_{a,b})(\kappa(i)) = \sigma(t_{a,b}(\kappa(i))) = \sigma\Bigg(\underbrace{(t_{a,b} \circ \kappa)}_{=\kappa}(i)\Bigg) = \sigma(\kappa(i)).$$

Hence, $a_{\kappa(i),(\Phi(\sigma))(\kappa(i))} = a_{\kappa(i),\sigma(\kappa(i))}$. This proves (1435).

Now,

$$\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}$$
$$=\underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma\text{ is even}}}}_{\substack{=\sum_{\sigma\in A_n}\\ \text{(since }A_n\text{ is the}\\ \text{set of all even}\\ \text{permutations}\\ \text{in }S_n)}}\underbrace{(-1)^\sigma}_{\substack{=1\\ \text{(since }\sigma\text{ is even)}}}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}+\underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma\text{ is odd}}}}_{\substack{=\sum_{\sigma\in C_n}\\ \text{(since }C_n\text{ is the}\\ \text{set of all odd}\\ \text{permutations}\\ \text{in }S_n)}}\underbrace{(-1)^\sigma}_{\substack{=-1\\ \text{(since }\sigma\text{ is odd)}}}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}$$
$$\text{(since every permutation }\sigma\in S_n\text{ is either even or odd, but not both)}$$
$$=\sum_{\sigma\in A_n}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}+\sum_{\sigma\in C_n}(-1)\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}$$
$$=\sum_{\sigma\in A_n}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}-\sum_{\sigma\in C_n}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}=0,$$

since

$$\sum_{\sigma\in C_n}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}=\sum_{\sigma\in A_n}\prod_{i=1}^{k}\underbrace{a_{\kappa(i),(\Phi(\sigma))(\kappa(i))}}_{\substack{=a_{\kappa(i),\sigma(\kappa(i))}\\ \text{(by (1435))}}}$$
$$\left(\begin{array}{c}\text{here, we have substituted }\Phi(\sigma)\text{ for }\sigma\text{ in}\\ \text{the sum, since the map }\Phi\text{ is a bijection}\end{array}\right)$$
$$=\sum_{\sigma\in A_n}\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}.$$

This proves Lemma 7.229. □

We can now solve part **(a)** of Exercise 6.54:

Proposition 7.230. Let $n\in\mathbb{N}$. Let $A=\left(a_{i,j}\right)_{1\le i\le n,\ 1\le j\le n}$ be an $n\times n$-matrix. Then,

$$\sum_{\sigma\in S_n}(-1)^\sigma\left(\sum_{i=1}^{n}a_{i,\sigma(i)}\right)^k=0$$

for each $k\in\{0,1,\ldots,n-2\}$.

Proof of Proposition 7.230. Let $k\in\{0,1,\ldots,n-2\}$. Then, each each map $\kappa:[k]\to[n]$ satisfies

$$\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^{k}a_{\kappa(i),\sigma(\kappa(i))}=0 \qquad (1436)$$

[640].

Now,

$$\sum_{\sigma\in S_n} (-1)^{\sigma} \underbrace{\left(\sum_{i=1}^{n} a_{i,\sigma(i)}\right)^{k}}_{\substack{=\sum_{\kappa:[k]\to[n]} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))} \\ \text{(by Corollary 7.228)}}}$$

$$= \sum_{\sigma\in S_n} (-1)^{\sigma} \sum_{\kappa:[k]\to[n]} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))} = \underbrace{\sum_{\sigma\in S_n} \sum_{\kappa:[k]\to[n]}}_{=\sum_{\kappa:[k]\to[n]} \sum_{\sigma\in S_n}} (-1)^{\sigma} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))}$$

$$= \sum_{\kappa:[k]\to[n]} \underbrace{\sum_{\sigma\in S_n} (-1)^{\sigma} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))}}_{\substack{=0 \\ \text{(by (1436))}}} = \sum_{\kappa:[k]\to[n]} 0 = 0.$$

This proves Proposition 7.230. □

Remark 7.231. Corollary 7.228, Lemma 7.229 and Proposition 7.230 all remain valid if the commutative ring $\mathbb{K}$ is replaced by a noncommutative ring $\mathbb{L}$, as long as we use the conventions made in Section 7.118. In fact, the proofs given above still work when $\mathbb{K}$ is replaced by $\mathbb{L}$, provided that we replace the reference to Lemma 6.22 by a reference to Lemma 7.220.

Part **(b)** of Exercise 6.54 is noticeably harder. We prepare to it by studying permutations in S_n:

Lemma 7.232. Let $n \in \mathbb{N}$ and $p \in [n]$. Let $\alpha \in S_n$ and $\beta \in S_n$. Assume that

$$\alpha(i) = \beta(i) \qquad \text{for each } i \in [n] \setminus \{p\}. \tag{1437}$$

Then, $\alpha = \beta$.

Proof of Lemma 7.232. We have $\alpha \in S_n$. In other words, α is a permutation of $[n]$ (since S_n is the set of all permutations of $\{1,2,\ldots,n\} = [n]$). In other words, α is a bijective map $[n] \to [n]$. The map α is bijective and thus injective. Hence, Lemma 1.3 **(c)** (applied to $U = [n]$, $V = [n]$, $f = \alpha$ and $S = [n] \setminus \{p\}$) shows that

$$|\alpha([n] \setminus \{p\})| = |[n] \setminus \{p\}| = |[n]| - 1 \qquad (\text{since } p \in [n]).$$

[640] *Proof of (1436):* Let $\kappa : [k] \to [n]$ be a map. Then, Lemma 1.3 **(a)** (applied to $U = [k]$, $V = [n]$, $f = \kappa$ and $S = [k]$) yields $|\kappa([k])| \le |[k]| = k \le n-2$ (since $k \in \{0,1,\ldots,n-2\}$), so that $|\kappa([k])| \le n-2 < n-1$. Hence, Lemma 7.229 yields $\sum_{\sigma\in S_n} (-1)^{\sigma} \prod_{i=1}^{k} a_{\kappa(i),\sigma(\kappa(i))} = 0$. This proves (1436).

Let G denote the subset $\alpha([n] \setminus \{p\})$ of $[n]$. Then, $G = \alpha([n] \setminus \{p\})$, and thus $|G| = |\alpha([n] \setminus \{p\})| = |[n]| - 1$.

But G is a subset of $[n]$; hence,

$$|[n] \setminus G| = |[n]| - \underbrace{|G|}_{=|[n]|-1} = |[n]| - (|[n]| - 1) = 1.$$

In other words, $[n] \setminus G$ is a 1-element set. Hence, $[n] \setminus G = \{q\}$ for some element q. Consider this q.

We have $q \in \{q\} = [n] \setminus G \subseteq [n]$.

Now, $\alpha(p) \notin G$ [641]. Combining this with $\alpha(p) \in [n]$, we obtain $\alpha(p) \in [n] \setminus G = \{q\}$. Thus, $\alpha(p) = q$.

Also, $\beta(p) \notin G$ [642]. Combining this with $\beta(p) \in [n]$, we obtain $\beta(p) \in [n] \setminus G = \{q\}$. Thus, $\beta(p) = q$. Comparing this with $\alpha(p) = q$, we obtain $\alpha(p) = \beta(p)$.

Now, each $i \in [n]$ satisfies $\alpha(i) = \beta(i)$ [643]. In other words, $\alpha = \beta$. This proves Lemma 7.232. □

Lemma 7.233. Let $n \geq 1$ be an integer. Then, the map

$$S_n \to \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\},$$
$$\tau \mapsto \tau \mid_{[n-1]}$$

is well-defined and bijective.

Proof of Lemma 7.233. We have $n - 1 \in \mathbb{N}$ (since $n \geq 1$) and thus $|[n-1]| = n - 1$.

Also, $n \in [n]$ (since $n \geq 1$) and $[n-1] = [n] \setminus \{n\} \subseteq [n]$.

Define a set Y by

$$Y = \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}. \tag{1438}$$

Each $\tau \in S_n$ satisfies $\tau \mid_{[n-1]} \in Y$ [644]. Hence, we can define a map $T : S_n \to Y$ by

$$\left(T(\tau) = \tau \mid_{[n-1]} \qquad \text{for all } \tau \in S_n\right).$$

[641] *Proof.* Assume the contrary. Thus, $\alpha(p) \in G$. Hence, $\alpha(p) \in G = \alpha([n] \setminus \{p\})$. In other words, there exists an $i \in [n] \setminus \{p\}$ satisfying $\alpha(p) = \alpha(i)$. Consider this i.

From $\alpha(p) = \alpha(i)$, we obtain $p = i$ (since α is injective). But $i \in [n] \setminus \{p\}$ shows that $i \notin \{p\}$, so that $i \neq p = i$. This is clearly a contradiction. This contradiction proves that our assumption was wrong. Qed.

[642] *Proof.* Assume the contrary. Thus, $\beta(p) \in G$. Hence, $\beta(p) \in G = \alpha([n] \setminus \{p\})$. In other words, there exists an $i \in [n] \setminus \{p\}$ satisfying $\beta(p) = \alpha(i)$. Consider this i.

From (1437), we obtain $\alpha(i) = \beta(i)$. Thus, $\beta(p) = \alpha(i) = \beta(i)$.

But recall that the map α is injective. Similarly, the map β is injective. Thus, from $\beta(p) = \beta(i)$, we obtain $p = i$. But $i \in [n] \setminus \{p\}$ shows that $i \notin \{p\}$, so that $i \neq p = i$. This is a contradiction. This contradiction proves that our assumption was wrong. Qed.

[643] *Proof.* Let $i \in [n]$. We must show that $\alpha(i) = \beta(i)$.

If $i = p$, then this is clearly true (because $\alpha(p) = \beta(p)$). Hence, for the rest of this proof, we WLOG assume that $i \neq p$. Hence, $i \in [n] \setminus \{p\}$. Thus, (1437) yields $\alpha(i) = \beta(i)$. Qed.

[644] *Proof.* Let $\tau \in S_n$.

Consider this map T.

Each $g \in Y$ satisfies $g \in T(S_n)$ [645]. In other words, $Y \subseteq T(S_n)$. Combining this with $T(S_n) \subseteq Y$ (which is obvious), we obtain $Y = T(S_n)$. In other words, the

---

We have $\tau \in S_n$. In other words, τ is a permutation of $[n]$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\} = [n]$). In other words, τ is a bijective map $[n] \to [n]$. The map τ is bijective and thus injective. Hence, Lemma 1.3 **(c)** (applied to $U = [n]$, $V = [n]$, $f = \tau$ and $S = [n-1]$) shows that $|\tau([n-1])| = |[n-1]| = n-1$.

Now, $\tau\mid_{[n-1]}$ is a map $[n-1] \to [n]$ and satisfies $\left| \underbrace{\left(\tau\mid_{[n-1]}\right)([n-1])}_{=\tau([n-1])} \right| = |\tau([n-1])| = n-1 \geq n-1$. Hence, $\tau\mid_{[n-1]}$ is a map $f : [n-1] \to [n]$ satisfying $|f([n-1])| \geq n-1$. In other words,

$$\tau\mid_{[n-1]} \in \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}.$$

In light of (1438), this rewrites as $\tau\mid_{[n-1]} \in Y$. Qed.

[645] *Proof.* Let $g \in Y$. Thus, $g \in Y = \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}$. In other words, g is a map $[n-1] \to [n]$ and satisfies $|g([n-1])| \geq n-1$.

Thus, $|g([n-1])| \geq n-1 = |[n-1]|$. Hence, Lemma 1.3 **(b)** (applied to $U = [n-1]$, $V = [n]$ and $f = g$) shows that the map g is injective. Hence, Lemma 1.3 **(c)** (applied to $U = [n-1]$, $V = [n]$, $f = g$ and $S = [n-1]$) shows that $|g([n-1])| = |[n-1]| = n-1$. Since $g([n-1])$ is a subset of $[n]$, we have

$$|[n] \setminus g([n-1])| = \underbrace{|[n]|}_{=n} - \underbrace{|g([n-1])|}_{=n-1} = n - (n-1) = 1.$$

In other words, $[n] \setminus g([n-1])$ is a 1-element set. In other words, $[n] \setminus g([n-1]) = \{p\}$ for some element p. Consider this p.

We have $p \in \{p\} = [n] \setminus g([n-1]) \subseteq [n]$.

Now, define a map $\sigma : [n] \to [n]$ by

$$\left(\sigma(i) = \begin{cases} g(i), & \text{if } i \in [n-1]; \\ p, & \text{if } i \notin [n-1] \end{cases} \qquad \text{for each } i \in [n] \right).$$

(This is well-defined, because each $i \in [n]$ satisfies $\begin{cases} g(i), & \text{if } i \in [n-1]; \\ p, & \text{if } i \notin [n-1] \end{cases} \in [n]$.)

We shall now show that the map σ is surjective.

Indeed, $n \notin \{1, 2, \ldots, n-1\} = [n-1]$ but $n \in [n]$. Hence, the definition of σ yields

$$\sigma(n) = \begin{cases} g(n), & \text{if } n \in [n-1]; \\ p, & \text{if } n \notin [n-1] \end{cases} = p \qquad (\text{since } n \notin [n-1]).$$

But $\sigma\left(\underbrace{n}_{\in[n]}\right) \in \sigma([n])$, so that $p = \sigma(n) \in \sigma([n])$.

Also, each $i \in [n-1]$ satisfies

$$\sigma(i) = \begin{cases} g(i), & \text{if } i \in [n-1]; \\ p, & \text{if } i \notin [n-1] \end{cases} = g(i) \qquad (\text{since } i \in [n-1]). \tag{1439}$$

map T is surjective.

If $\alpha \in S_n$ and $\beta \in S_n$ satisfy $T(\alpha) = T(\beta)$, then $\alpha = \beta$ [646]. In other words, the map T is injective.

The map T is surjective and injective. In other words, the map T is bijective.

Recall that T is a map from S_n to Y. In other words, T is a map from S_n to $\{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}$ (since $Y = \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}$). Hence, the map

$$S_n \to \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}, \\ \tau \mapsto \tau \mid_{[n-1]}$$

Now,

$$g([n-1]) = \left\{ \underbrace{g(i)}_{\substack{=\sigma(i) \\ \text{(by (1439))}}} \mid i \in [n-1] \right\} = \{\sigma(i) \mid i \in [n-1]\} = \sigma\left(\underbrace{[n-1]}_{\subseteq [n]}\right) \subseteq \sigma([n]).$$

If X is a set, and if Y is a subset of X, then $X = Y \cup (X \setminus Y)$. Applying this to $X = [n]$ and $Y = g([n-1])$, we obtain

$$[n] = \underbrace{g([n-1])}_{\subseteq \sigma([n])} \cup \underbrace{([n] \setminus g([n-1]))}_{\substack{=\{p\} \subseteq \sigma([n]) \\ \text{(since } p \in \sigma([n]))}} \subseteq \sigma([n]) \cup \sigma([n]) = \sigma([n]).$$

Combining this with the obvious relation $\sigma([n]) \subseteq [n]$, we obtain $[n] = \sigma([n])$. In other words, the map σ is surjective.

Lemma 1.4 (applied to $U = [n]$, $V = [n]$ and $f = \sigma$) shows that we have the following logical equivalence:

$$(\sigma \text{ is surjective}) \iff (\sigma \text{ is bijective})$$

(since $|[n]| \leq |[n]|$). Thus, σ is bijective (since σ is surjective). Hence, σ is a bijective map $[n] \to [n]$. In other words, σ is a permutation of $[n]$. In other words, $\sigma \in S_n$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\} = [n]$). Hence, $T(\sigma)$ is well-defined.

The definition of T yields $T(\sigma) = \sigma \mid_{[n-1]}$. Thus, each $i \in [n-1]$ satisfies

$$(T(\sigma))(i) = \left(\sigma \mid_{[n-1]}\right)(i) = \sigma(i) = g(i) \qquad \text{(by (1439))}.$$

In other words, $T(\sigma) = g$. Thus, $g = T\left(\underbrace{\sigma}_{\in S_n}\right) \in T(S_n)$. Qed.

[646] *Proof.* Let $\alpha \in S_n$ and $\beta \in S_n$ be such that $T(\alpha) = T(\beta)$. We must show that $\alpha = \beta$.

The definition of T yields $T(\alpha) = \alpha \mid_{[n-1]}$ and $T(\beta) = \beta \mid_{[n-1]}$. Hence, $\alpha \mid_{[n-1]} = T(\alpha) = T(\beta) = \beta \mid_{[n-1]}$.

Now, each $i \in [n-1]$ satisfies $\alpha(i) = \underbrace{\left(\alpha \mid_{[n-1]}\right)}_{=\beta|_{[n-1]}}(i) = \left(\beta \mid_{[n-1]}\right)(i) = \beta(i)$. Thus, $\alpha(i) = \beta(i)$ for each $i \in [n-1]$. In other words, $\alpha(i) = \beta(i)$ for each $i \in [n] \setminus \{n\}$ (since $[n-1] = [n] \setminus \{n\}$). Thus, Lemma 7.232 (applied to $p = n$) yields $\alpha = \beta$. Qed.

is precisely the map T (since $T(\tau) = \tau\mid_{[n-1]}$ for all $\tau \in S_n$). Hence, this map is well-defined and bijective (since T is bijective). This proves Lemma 7.233. $\square$

Lemma 7.234. Let $n \geq 1$ be an integer. Let $\tau \in S_n$. Then,

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n-1} a_{\left(\tau|_{[n-1]}\right)(i),\sigma\left(\left(\tau|_{[n-1]}\right)(i)\right)} = \sum_{q=1}^{n} (-1)^{\tau(n)+q} \det\left(A_{\sim(\tau(n)),\sim q}\right).$$

Proof of Lemma 7.234. We have $\tau \in S_n$. In other words, τ is a permutation of $[n]$ (since S_n is the set of all permutations of $\{1, 2, \ldots, n\} = [n]$). In other words, τ is a bijective map $[n] \to [n]$. Thus, $\tau : [n] \to [n]$ is a bijection.

Also, $n \in [n]$ (since $n \geq 1$) and $[n-1] = [n] \setminus \{n\} \subseteq [n]$.

Let $p = \tau(n)$. Then, $p = \tau(n) \in [n] = \{1, 2, \ldots, n\}$. From $p = \tau(n)$, we obtain $n = \tau^{-1}(p)$.

If $b_1, b_2, \ldots, b_n$ are n elements of $\mathbb{K}$, then

$$\prod_{i=1}^{n-1} b_{\left(\tau|_{[n-1]}\right)(i)} = \prod_{\substack{i \in \{1,2,\ldots,n\};\\ i \neq p}} b_i \qquad (1440)$$

[647].

[647] *Proof of (1440):* Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$.

For each $i \in [n]$, we have the following chain of logical equivalences:

$$\left(i \neq \underbrace{n}_{=\tau^{-1}(p)}\right) \iff \left(i \neq \tau^{-1}(p)\right) \iff (\tau(i) \neq p)$$

(since τ is a bijection). In other words, for each $i \in [n]$, the condition $(i \neq n)$ holds if and only if the condition $(\tau(i) \neq p)$ holds. Thus, we have the following equality of product signs:

$$\prod_{\substack{i \in [n];\\ i \neq n}} = \prod_{\substack{i \in [n];\\ \tau(i) \neq p}}.$$

Hence,

$$\prod_{\substack{i \in [n];\\ \tau(i) \neq p}} = \prod_{\substack{i \in [n];\\ i \neq n}} = \prod_{i \in [n] \setminus \{n\}} = \prod_{i \in [n-1]} \qquad (\text{since } [n] \setminus \{n\} = [n-1])$$
$$= \prod_{i=1}^{n-1}. \qquad (1441)$$

Now, the map $\tau : [n] \to [n]$ is a bijection. Thus, we can substitute $\tau(i)$ for i in the product

We have

$$\underbrace{\sum_{\sigma\in S_n}}_{\substack{=\sum\limits_{q\in[n]}\sum\limits_{\substack{\sigma\in S_n;\\ \sigma(p)=q}} \\ \text{(because for each } \sigma\in S_n\text{, there} \\ \text{exists a unique } q\in[n] \\ \text{such that } \sigma(p)=q\text{)}}} (-1)^{\sigma} \underbrace{\prod_{i=1}^{n-1} a_{\left(\tau|_{[n-1]}\right)(i),\sigma\left(\left(\tau|_{[n-1]}\right)(i)\right)}}_{\substack{=\prod\limits_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} a_{i,\sigma(i)} \\ \text{(by (1440), applied to } b_k=a_{k,\sigma(k)}\text{)}}}$$

$$= \underbrace{\sum_{q\in[n]}}_{=\sum_{q=1}^{n}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ \sigma(p)=q}} (-1)^{\sigma} \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} a_{i,\sigma(i)}}_{\substack{=(-1)^{p+q}\det\left(A_{\sim p,\sim q}\right) \\ \text{(by Lemma 6.84)}}}$$

$$= \sum_{q=1}^{n} (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right) = \sum_{q=1}^{n} (-1)^{\tau(n)+q} \det\left(A_{\sim(\tau(n)),\sim q}\right)$$

(since $p = \tau(n)$). This proves Lemma 7.234. □

Lemma 7.235. Let $n \geq 1$ be an integer.

(a) We have $|\{\tau \in S_n \mid \tau(n) = n\}| = (n-1)!$.

(b) Let $p \in [n]$ and $q \in [n]$. Then,

$$|\{\tau \in S_n \mid \tau(p) = q\}| = (n-1)!.$$

Proof of Lemma 7.235. We have $n \geq 1$. Thus, $n - 1 \in \mathbb{N}$.

$\prod\limits_{\substack{i\in[n];\\ i\neq p}} b_i$. We thus obtain

$$\prod_{\substack{i\in[n];\\ i\neq p}} b_i = \underbrace{\prod_{\substack{i\in[n];\\ \tau(i)\neq p}}}_{\substack{=\prod_{i=1}^{n-1} \\ \text{(by (1441))}}} b_{\tau(i)} = \prod_{i=1}^{n-1} b_{\tau(i)}. \tag{1442}$$

Now,

$$\prod_{i=1}^{n-1} \underbrace{b_{\left(\tau|_{[n-1]}\right)(i)}}_{\substack{=b_{\tau(i)} \\ \text{(since } \left(\tau|_{[n-1]}\right)(i)=\tau(i)\text{)}}} = \prod_{i=1}^{n-1} b_{\tau(i)} = \prod_{\substack{i\in[n];\\ i\neq p}} b_i \qquad \text{(by (1442))}$$

$$= \prod_{\substack{i\in\{1,2,\ldots,n\};\\ i\neq p}} b_i \qquad \left(\text{since } [n] = \{1,2,\ldots,n\}\right).$$

This proves (1440).

Hence, Corollary 7.81 (applied to $n-1$ instead of n) yields $|S_{n-1}| = (n-1)!$.

(a) Clearly, n is a positive integer (since n is an integer satisfying $n \geq 1$).

Define a subset T of S_n by $T = \{\tau \in S_n \mid \tau(n) = n\}$. Define a map $\Phi : S_{n-1} \to T$ as in the proof of Lemma 6.44. Then, the map Φ is a bijection[648]. Hence, there exists a bijection $S_{n-1} \to T$ (namely, Φ). Thus, $|T| = |S_{n-1}| = (n-1)!$. Since $T = \{\tau \in S_n \mid \tau(n) = n\}$, this rewrites as $|\{\tau \in S_n \mid \tau(n) = n\}| = (n-1)!$. This proves Lemma 7.235 **(a)**.

(b) Define a map $T : \{\tau \in S_n \mid \tau(n) = n\} \to \{\tau \in S_n \mid \tau(p) = q\}$ as in Lemma 6.83 **(d)**. Then, Lemma 6.83 **(d)** shows that this map T is well-defined and bijective. Thus, there exists a bijection from $\{\tau \in S_n \mid \tau(n) = n\}$ to $\{\tau \in S_n \mid \tau(p) = q\}$ (namely, T). Hence,

$$|\{\tau \in S_n \mid \tau(p) = q\}| = |\{\tau \in S_n \mid \tau(n) = n\}| = (n-1)!$$

(by Lemma 7.235 **(a)**). This proves Lemma 7.235 **(b)**. □

Finally, we are approaching part **(b)** of Exercise 6.54:

Proposition 7.236. Let $n \geq 1$ be an integer. Let $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq n}$ be an $n \times n$-matrix. Then,

$$\sum_{\sigma \in S_n} (-1)^{\sigma} \left(\sum_{i=1}^{n} a_{i,\sigma(i)} \right)^{n-1} = (n-1)! \cdot \sum_{p=1}^{n} \sum_{q=1}^{n} (-1)^{p+q} \det \left(A_{\sim p, \sim q} \right).$$

Proof of Proposition 7.236. From $n \geq 1$, we obtain $n-1 \in \mathbb{N}$. Hence, for every $\sigma \in S_n$, we have

$$\left(\sum_{i=1}^{n} a_{i,\sigma(i)} \right)^{n-1} = \sum_{\kappa : [n-1] \to [n]} \prod_{i=1}^{n-1} a_{\kappa(i), \sigma(\kappa(i))} \tag{1443}$$

(by Corollary 7.228, applied to $k = n-1$).

Lemma 7.233 shows that the map

$$S_n \to \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\},$$
$$\tau \mapsto \tau \mid_{[n-1]}$$

is well-defined and bijective. Thus, this map is a bijection.

We have $[n] = \{1, 2, \ldots, n\}$. Hence, we have the following equality of summation signs:

$$\sum_{p \in [n]} = \sum_{p \in \{1,2,\ldots,n\}} = \sum_{p=1}^{n}. \tag{1444}$$

[648] This was shown in the proof of Lemma 6.44.

Now,

$$\begin{aligned}
&\sum_{\sigma\in S_n}(-1)^{\sigma}\underbrace{\left(\sum_{i=1}^{n}a_{i,\sigma(i)}\right)^{n-1}}_{\substack{=\sum_{\kappa:[n-1]\to[n]}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}\\ \text{(by (1443))}}}\\
&=\sum_{\sigma\in S_n}(-1)^{\sigma}\sum_{\kappa:[n-1]\to[n]}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}=\underbrace{\sum_{\sigma\in S_n}\sum_{\kappa:[n-1]\to[n]}}_{=\sum_{\kappa:[n-1]\to[n]}\sum_{\sigma\in S_n}}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}\\
&=\sum_{\kappa:[n-1]\to[n]}\sum_{\sigma\in S_n}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}\\
&=\sum_{\substack{\kappa:[n-1]\to[n];\\ |\kappa([n-1])|<n-1}}\underbrace{\sum_{\sigma\in S_n}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}}_{\substack{=0\\ \text{(by Lemma 7.229 (applied to } k=n-1))}}\\
&\qquad+\underbrace{\sum_{\substack{\kappa:[n-1]\to[n];\\ |\kappa([n-1])|\geq n-1}}}_{=\sum_{\kappa\in\{f:[n-1]\to[n]\ \mid\ |f([n-1])|\geq n-1\}}}\sum_{\sigma\in S_n}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}\\
&\qquad\left(\begin{array}{c}\text{since each map } \kappa:[n-1]\to[n] \text{ satisfies either } |\kappa([n-1])|<n-1\\ \text{or } |\kappa([n-1])|\geq n-1 \text{ (but not both)}\end{array}\right)\\
&=\underbrace{\sum_{\substack{\kappa:[n-1]\to[n];\\ |\kappa([n-1])|<n-1}}0}_{=0}+\sum_{\kappa\in\{f:[n-1]\to[n]\ \mid\ |f([n-1])|\geq n-1\}}\sum_{\sigma\in S_n}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}\\
&=\sum_{\kappa\in\{f:[n-1]\to[n]\ \mid\ |f([n-1])|\geq n-1\}}\sum_{\sigma\in S_n}(-1)^{\sigma}\prod_{i=1}^{n-1}a_{\kappa(i),\sigma(\kappa(i))}
\end{aligned}$$

$$= \sum_{\tau \in S_n} \underbrace{\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n-1} a_{\left(\tau|_{[n-1]}\right)(i), \sigma\left(\left(\tau|_{[n-1]}\right)(i)\right)}}_{\substack{= \sum_{q=1}^{n} (-1)^{\tau(n)+q} \det\left(A_{\sim(\tau(n)),\sim q}\right) \\ \text{(by Lemma 7.234)}}}$$

$$\left(\begin{array}{c} \text{here, we have substituted } \tau\mid_{[n-1]} \text{ for } \kappa \text{ in the outer sum,} \\ \text{since the} \\ \text{map } S_n \to \{f : [n-1] \to [n] \mid |f([n-1])| \geq n-1\}, \ \tau \mapsto \tau\mid_{[n-1]} \\ \text{is a bijection} \end{array} \right)$$

$$= \underbrace{\sum_{\tau \in S_n}}_{\substack{= \sum_{p \in [n]} \sum_{\substack{\tau \in S_n; \\ \tau(n) = p}} \\ \text{(because for each } \tau \in S_n \text{, there} \\ \text{exists a unique } p \in [n] \\ \text{such that } \tau(n) = p)}} \sum_{q=1}^{n} (-1)^{\tau(n)+q} \det\left(A_{\sim(\tau(n)),\sim q}\right)$$

$$= \sum_{p \in [n]} \underbrace{\sum_{\substack{\tau \in S_n; \\ \tau(n) = p}} \sum_{q=1}^{n}}_{= \sum_{q=1}^{n} \sum_{\substack{\tau \in S_n; \\ \tau(n) = p}}} \underbrace{(-1)^{\tau(n)+q} \det\left(A_{\sim(\tau(n)),\sim q}\right)}_{\substack{= (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right) \\ \text{(since } \tau(n) = p)}}$$

$$= \sum_{p \in [n]} \sum_{q=1}^{n} \underbrace{\sum_{\substack{\tau \in S_n; \\ \tau(n) = p}} (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right)}_{= |\{\tau \in S_n \mid \tau(n) = p\}| (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right)}$$

$$= \underbrace{\sum_{p \in [n]}}_{\substack{= \sum_{p=1}^{n} \\ \text{(by (1444))}}} \sum_{q=1}^{n} \underbrace{|\{\tau \in S_n \mid \tau(n) = p\}|}_{\substack{= (n-1)! \\ \text{(by Lemma 7.235 (b)} \\ \text{(applied to } n \text{ and } p \text{ instead of } p \text{ and } q))}} (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right)$$

$$= \sum_{p=1}^{n} \sum_{q=1}^{n} (n-1)! (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right) = (n-1)! \cdot \sum_{p=1}^{n} \sum_{q=1}^{n} (-1)^{p+q} \det\left(A_{\sim p,\sim q}\right).$$

This proves Proposition 7.236. □

Solution to Exercise 6.54. Exercise 6.54 **(a)** follows from Proposition 7.230.

Exercise 6.54 **(b)** follows from Proposition 7.236. □

7.123. Solution to Exercise 6.55

7.123.1. Solving the exercise

We prepare for the solution to Exercise 6.55 by showing a few lemmas.

Lemma 7.237. Let $n \in \mathbb{N}$. We let $\mathbf{E}$ be the subset

$$\{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct}\}$$

of $\mathbb{N}^n$. We let $\mathbf{I}$ be the subset

$$\{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n \mid k_1 < k_2 < \cdots < k_n\}$$

of $\mathbb{N}^n$. Then, the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

is well-defined and is a bijection.

Lemma 7.237 is identical to Lemma 6.41 except that it uses the set $\mathbb{N}$ instead of the set $[m]$ used in the latter lemma. Thus, it should not be a surprise that its proof is the same as the proof of the latter lemma:

Proof of Lemma 7.237. In order to obtain a proof of Lemma 7.237, it is sufficient to replace every appearance of "$[m]$" by "$\mathbb{N}$" in the proof of Lemma 6.41. (Of course, the notation $[n]$ should be understood to mean the set $\{1, 2, \ldots, n\}$ in this proof.) □

Lemma 7.238. Let $n \in \mathbb{N}$. Let $(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$ be such that $k_1 < k_2 < \cdots < k_n$.

(a) We have $k_1 + k_2 + \cdots + k_n \geq \dbinom{n}{2}$.

(b) If $k_1 + k_2 + \cdots + k_n \leq \dbinom{n}{2}$, then $(k_1, k_2, \ldots, k_n) = (0, 1, \ldots, n-1)$.

Proof of Lemma 7.238. **(b)** Assume that $k_1 + k_2 + \cdots + k_n \leq \dbinom{n}{2}$. We must prove that $(k_1, k_2, \ldots, k_n) = (0, 1, \ldots, n-1)$.

For each $i \in \{1, 2, \ldots, n\}$, define an integer z_i by $z_i = k_i - i$.

For each $i \in \{1, 2, \ldots, n-1\}$, we have $z_i \leq z_{i+1}$ [649]. In other words, $z_1 \leq z_2 \leq \cdots \leq z_n$.

[649] *Proof.* Let $i \in \{1, 2, \ldots, n-1\}$. The definition of z_i yields $z_i = k_i - i$. The definition of z_{i+1} yields $z_{i+1} = k_{i+1} - (i+1)$. But $k_i < k_{i+1}$ (since $k_1 < k_2 < \cdots < k_n$), and thus $k_i \leq k_{i+1} - 1$ (since k_i and k_{i+1} are integers). Now, $z_i = \underbrace{k_i}_{\leq k_{i+1}-1} - i \leq k_{i+1} - 1 - i = k_{i+1} - (i+1) = z_{i+1}$. Qed.

For each $j \in \{1, 2, \ldots, n\}$, we have

$$k_j \geq j - 1 \tag{1445}$$

[650].

Now, each $j \in \{1, 2, \ldots, n\}$ satisfies

$$k_j = j - 1. \tag{1446}$$

[*Proof of (1446):* Let $j \in \{1, 2, \ldots, n\}$. We must prove that $k_j = j - 1$.

Assume the contrary. Thus, $k_j \neq j - 1$. Combining this with (1445), we obtain $k_j > j - 1$.

Now, $k_1 + k_2 + \cdots + k_n \leq \dbinom{n}{2}$, so that

$$\begin{aligned}
\dbinom{n}{2} &\geq k_1 + k_2 + \cdots + k_n = \sum_{i \in \{1,2,\ldots,n\}} k_i \\
&= \underbrace{k_j}_{> j-1} + \sum_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq j}} \underbrace{k_i}_{\substack{\geq i-1 \\ \text{(by (1445) (applied} \\ \text{to } i \text{ instead of } j))}} \qquad \left(\begin{array}{c}\text{here, we have split off the addend} \\ \text{for } i = j \text{ from the sum}\end{array}\right) \\
&> (j-1) + \sum_{\substack{i \in \{1,2,\ldots,n\}; \\ i \neq j}} (i-1) = \sum_{i \in \{1,2,\ldots,n\}} (i-1) \\
&= 0 + 1 + \cdots + (n-1) = \sum_{r=0}^{n-1} r = \dbinom{n}{2}
\end{aligned}$$

(by Lemma 7.184). This is absurd. Thus, we have found a contradiction. This contradiction shows that our assumption was wrong. Hence, $k_j = j - 1$ is proven. This proves (1446).]

Now, (1446) shows that $(k_1, k_2, \ldots, k_n) = (1-1, 2-1, \ldots, n-1) = (0, 1, \ldots, n-1)$. This proves Lemma 7.238 **(b)**.

(a) Assume the contrary. Thus, $k_1 + k_2 + \cdots + k_n < \dbinom{n}{2}$. Hence, $k_1 + k_2 + \cdots + k_n \leq \dbinom{n}{2}$. Therefore, Lemma 7.238 **(b)** shows that $(k_1, k_2, \ldots, k_n) = (0, 1, \ldots, n-1)$.

[650]*Proof of (1445):* Let $j \in \{1, 2, \ldots, n\}$. Thus, $1 \leq j \leq n$. Hence, $n \geq 1$, so that $1 \in \{1, 2, \ldots, n\}$.

But recall that $z_1 \leq z_2 \leq \cdots \leq z_n$. In other words, we have $z_u \leq z_v$ whenever u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u \leq v$. Applying this to $u = 1$ and $v = j$, we obtain $z_1 \leq z_j$ (since $1 \leq j$). But the definition of z_1 yields $z_1 = \underbrace{k_1}_{\substack{\geq 0 \\ (\text{since } k_1 \in \mathbb{N})}} - 1 \geq 0 - 1 = -1$. Hence, $-1 \leq z_1 \leq z_j = k_j - j$ (by the definition of z_j). Hence, $k_j - j \geq -1$, so that $k_j \geq j - 1$. This proves (1445).

Thus,

$$k_1 + k_2 + \cdots + k_n = 0 + 1 + \cdots + (n-1) = \sum_{r=0}^{n-1} r = \binom{n}{2}$$

(by Lemma 7.184), which contradicts $k_1 + k_2 + \cdots + k_n < \binom{n}{2}$. This contradiction shows that our assumption was wrong. Hence, Lemma 7.238 **(a)** is proven. □

Lemma 7.239. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$.

Let $(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$.

(a) We have

$$\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} \left(a_i b_{\sigma(i)} \right)^{k_i} = \left(\prod_{i=1}^{n} a_i^{k_i} \right) \det \left(\left(b_i^{k_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right).$$

(b) If the integers $k_1, k_2, \ldots, k_n$ are not distinct, then

$$\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} \left(a_i b_{\sigma(i)} \right)^{k_i} = 0.$$

Proof of Lemma 7.239. **(a)** Define an $n \times n$-matrix A by $A = \left(b_j^{k_i} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Then, the equality (341) (applied to $b_j^{k_i}$ instead of $a_{i,j}$) yields

$$\det A = \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} b_{\sigma(i)}^{k_i}.$$

But $A = \left(b_j^{k_i} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus, $A^T = \left(b_i^{k_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of the transpose matrix A^T). Exercise 6.4 now yields

$$\det \left(A^T \right) = \det A = \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} b_{\sigma(i)}^{k_i}. \tag{1447}$$

Now,

$$\begin{aligned}
&\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^n\underbrace{\left(a_ib_{\sigma(i)}\right)^{k_i}}_{=a_i^{k_i}b_{\sigma(i)}^{k_i}} \\
&=\sum_{\sigma\in S_n}(-1)^\sigma\underbrace{\prod_{i=1}^n\left(a_i^{k_i}b_{\sigma(i)}^{k_i}\right)}_{=\left(\prod_{i=1}^n a_i^{k_i}\right)\left(\prod_{i=1}^n b_{\sigma(i)}^{k_i}\right)}=\sum_{\sigma\in S_n}(-1)^\sigma\left(\prod_{i=1}^n a_i^{k_i}\right)\left(\prod_{i=1}^n b_{\sigma(i)}^{k_i}\right) \\
&=\left(\prod_{i=1}^n a_i^{k_i}\right)\underbrace{\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^n b_{\sigma(i)}^{k_i}}_{\substack{=\det\left(A^T\right)\\ \text{(by (1447))}}}=\left(\prod_{i=1}^n a_i^{k_i}\right)\det\left(\underbrace{A^T}_{=\left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}}\right) \\
&=\left(\prod_{i=1}^n a_i^{k_i}\right)\det\left(\left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).
\end{aligned}$$

This proves Lemma 7.239 **(a)**.

(b) Assume that the integers $k_1, k_2, \ldots, k_n$ are not distinct. Thus, there exist two distinct elements p and q of $\{1, 2, \ldots, n\}$ such that $k_p = k_q$. Consider these p and q. Now,

$$\begin{aligned}
&\left(\text{the } p\text{-th column of the matrix } \left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \\
&=\left(b_i^{k_p}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}=\left(b_i^{k_q}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\qquad\left(\text{since } k_p=k_q\right) \\
&=\left(\text{the } q\text{-th column of the matrix } \left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right).
\end{aligned}$$

Thus, the matrix $\left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ has two equal columns (namely, the p-th column and the q-th column), because p and q are distinct. Therefore, Exercise 6.7 **(f)** (applied to $\left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ instead of A) yields $\det\left(\left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)=0$.

Now, Lemma 7.239 **(a)** yields

$$\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}=\left(\prod_{i=1}^n a_i^{k_i}\right)\underbrace{\det\left(\left(b_i^{k_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)}_{=0}=0.$$

This proves Lemma 7.239 **(b)**. □

We shall use the notation introduced in Exercise 6.2 from now on until the end of the present section.

Lemma 7.240. Let $n \in \mathbb{N}$. Let $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ and $\sigma \in S_n$.

(a) We have $g_{\sigma(1)} + g_{\sigma(2)} + \cdots + g_{\sigma(n)} = g_1 + g_2 + \cdots + g_n$.

(b) We have $\mathbf{m}\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right) = \mathbf{m}(g_1, g_2, \ldots, g_n)$.

Proof of Lemma 7.240. We have $\sigma \in S_n$. In other words, σ is a permutation of the set $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$). In other words, σ is a bijection $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. Hence, we can substitute $\sigma(i)$ for i in the sum $\sum\limits_{i \in \{1,2,\ldots,n\}} g_i$. We thus obtain

$$\sum_{i \in \{1,2,\ldots,n\}} g_i = \sum_{i \in \{1,2,\ldots,n\}} g_{\sigma(i)} = g_{\sigma(1)} + g_{\sigma(2)} + \cdots + g_{\sigma(n)}.$$

Hence,

$$g_{\sigma(1)} + g_{\sigma(2)} + \cdots + g_{\sigma(n)} = \sum_{i \in \{1,2,\ldots,n\}} g_i = g_1 + g_2 + \cdots + g_n. \tag{1448}$$

This proves Lemma 7.240 **(a)**.

(b) We have proven that $g_{\sigma(1)} + g_{\sigma(2)} + \cdots + g_{\sigma(n)} = g_1 + g_2 + \cdots + g_n$. A similar argument shows that

$$g_{\sigma(1)}! g_{\sigma(2)}! \cdots g_{\sigma(n)}! = g_1! g_2! \cdots g_n!. \tag{1449}$$

The definition of $\mathbf{m}(g_1, g_2, \ldots, g_n)$ yields

$$\mathbf{m}(g_1, g_2, \ldots, g_n) = \frac{(g_1 + g_2 + \cdots + g_n)!}{g_1! g_2! \cdots g_n!}.$$

But the definition of $\mathbf{m}\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$ yields

$$\begin{aligned} \mathbf{m}\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right) &= \frac{\left(g_{\sigma(1)} + g_{\sigma(2)} + \cdots + g_{\sigma(n)}\right)!}{g_{\sigma(1)}! g_{\sigma(2)}! \cdots g_{\sigma(n)}!} \\ &= \frac{(g_1 + g_2 + \cdots + g_n)!}{g_1! g_2! \cdots g_n!} \qquad \text{(by (1448) and (1449))} \\ &= \mathbf{m}(g_1, g_2, \ldots, g_n). \end{aligned}$$

This proves Lemma 7.240 **(b)**. ☐

Now, we can attack part **(c)** of Exercise 6.55:

Lemma 7.241. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$.

Let $k \in \mathbb{N}$. Then,

$$\begin{aligned}
&\sum_{\sigma \in S_n} (-1)^\sigma \left(\sum_{i=1}^n a_i b_{\sigma(i)} \right)^k \\
&= \sum_{\substack{(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n; \\ g_1 < g_2 < \cdots < g_n; \\ g_1 + g_2 + \cdots + g_n = k}} \mathbf{m}(g_1, g_2, \ldots, g_n) \\
&\qquad \cdot \det \left(\left(a_i^{g_j} \right)_{1 \le i \le n, \ 1 \le j \le n} \right) \cdot \det \left(\left(b_i^{g_j} \right)_{1 \le i \le n, \ 1 \le j \le n} \right).
\end{aligned}$$

Here, we are using the notation introduced in Exercise 6.2.

Proof of Lemma 7.241. We shall use the notations of Lemma 7.237. Recall that

$$\begin{aligned}
\mathbf{E} &= \{ (k_1, k_2, \ldots, k_n) \in \mathbb{N}^n \mid \text{ the integers } k_1, k_2, \ldots, k_n \text{ are distinct} \} \\
&= \{ (g_1, g_2, \ldots, g_n) \in \mathbb{N}^n \mid \text{ the integers } g_1, g_2, \ldots, g_n \text{ are distinct} \} \qquad (1450)
\end{aligned}$$

(here, we renamed the index $(k_1, k_2, \ldots, k_n)$ as $(g_1, g_2, \ldots, g_n)$) and

$$\begin{aligned}
\mathbf{I} &= \{ (k_1, k_2, \ldots, k_n) \in \mathbb{N}^n \mid k_1 < k_2 < \cdots < k_n \} \\
&= \{ (g_1, g_2, \ldots, g_n) \in \mathbb{N}^n \mid g_1 < g_2 < \cdots < g_n \} \qquad (1451)
\end{aligned}$$

(here, we renamed the index $(k_1, k_2, \ldots, k_n)$ as $(g_1, g_2, \ldots, g_n)$).

Lemma 7.237 says that the map

$$\begin{aligned}
\mathbf{I} \times S_n &\to \mathbf{E}, \\
((g_1, g_2, \ldots, g_n), \sigma) &\mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)} \right)
\end{aligned}$$

is well-defined and is a bijection.

Next, let us make three observations:

- Every $(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$ satisfying $(k_1, k_2, \ldots, k_n) \notin \mathbf{E}$ satisfies

$$\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^n \left(a_i b_{\sigma(i)} \right)^{k_i} = 0 \qquad (1452)$$

[651].

[651] *Proof of (1452):* Let $(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$ be such that $(k_1, k_2, \ldots, k_n) \notin \mathbf{E}$.

Let us first show that the integers $k_1, k_2, \ldots, k_n$ are not distinct. Indeed, we assume the contrary. Thus, the integers $k_1, k_2, \ldots, k_n$ are distinct. Hence, $(k_1, k_2, \ldots, k_n)$ is an element of $\mathbb{N}^n$ having the property that the integers $k_1, k_2, \ldots, k_n$ are distinct. In other words, $(k_1, k_2, \ldots, k_n)$ is

- Every $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ and every $\sigma \in S_n$ satisfy

$$\det\left(\left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}\right) = (-1)^\sigma \cdot \det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right) \qquad (1453)$$

[652].

---

an element $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ such that the integers $g_1, g_2, \ldots, g_n$ are distinct. In other words,

$$(k_1, k_2, \ldots, k_n) \in \{(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n \mid \text{the integers } g_1, g_2, \ldots, g_n \text{ are distinct}\}.$$

In light of (1450), this rewrites as $(k_1, k_2, \ldots, k_n) \in \mathbf{E}$. This contradicts $(k_1, k_2, \ldots, k_n) \notin \mathbf{E}$. This contradiction proves that our assumption was wrong. Hence, we have proven that the integers $k_1, k_2, \ldots, k_n$ are not distinct. Hence, Lemma 7.239 **(b)** yields $\sum\limits_{\sigma\in S_n} (-1)^\sigma \prod\limits_{i=1}^n \left(a_i b_{\sigma(i)}\right)^{k_i} = 0$. This proves (1452).

[652]*Proof of (1453):* Let $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ and $\sigma \in S_n$. Then, $g_1, g_2, \ldots, g_n$ are elements of $\mathbb{N}$ (since $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$). Hence, $g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}$ are elements of $\mathbb{N}$. In other words, $\left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right) \in \mathbb{N}^n$.

Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Recall that $\sigma \in S_n$. In other words, σ is a permutation of the set $\{1, 2, \ldots, n\}$ (since S_n is the set of all permutations of the set $\{1, 2, \ldots, n\}$). In other words, σ is a permutation of the set $[n]$ (since $\{1, 2, \ldots, n\} = [n]$). In other words, σ is a bijective map $[n] \to [n]$.

Define an $n \times n$-matrix A by $A = \left(b_j^{g_i}\right)_{1\le i\le n,\ 1\le j\le n}$. Define an $n \times n$-matrix A_σ by $A_\sigma = \left(b_j^{g_{\sigma(i)}}\right)_{1\le i\le n,\ 1\le j\le n}$. Lemma 6.17 **(a)** (applied to σ, A, $b_j^{g_i}$ and A_σ instead of κ, B, $b_{i,j}$ and B_κ) then yields $\det(A_\sigma) = (-1)^\sigma \cdot \det A$.

But $A = \left(b_j^{g_i}\right)_{1\le i\le n,\ 1\le j\le n}$. Thus, $A^T = \left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}$ (by the definition of the transpose matrix A^T). Exercise 6.4 now yields $\det\left(A^T\right) = \det A$. Hence,

$$\det A = \det\left(\underbrace{A^T}_{=\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}}\right) = \det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right).$$

On the other hand, $A_\sigma = \left(b_j^{g_{\sigma(i)}}\right)_{1\le i\le n,\ 1\le j\le n}$. Hence, $(A_\sigma)^T = \left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}$ (by the definition of the transpose matrix $(A_\sigma)^T$). Exercise 6.4 (applied to A_σ instead of A) now yields $\det\left((A_\sigma)^T\right) = \det(A_\sigma)$. Since $(A_\sigma)^T = \left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}$, this rewrites as

$$\begin{aligned}\det\left(\left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}\right) &= \det(A_\sigma) = (-1)^\sigma \cdot \underbrace{\det A}_{=\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)} \\ &= (-1)^\sigma \cdot \det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right).\end{aligned}$$

This proves (1453).

- Every $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfies

$$\det\left(\left(a_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \sum_{\sigma\in S_n}\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)(-1)^\sigma \qquad (1454)$$

[653].

Now, each $\sigma \in S_n$ satisfies

$$\left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^k = \sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}} \mathbf{m}(k_1, k_2, \ldots, k_n) \prod_{i=1}^n \left(a_i b_{\sigma(i)}\right)^{k_i} \qquad (1455)$$

[654].

[653] *Proof of (1454):* Let $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$. The equality (341) (applied to $\left(a_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$ and $a_i^{g_j}$ instead of A and $a_{i,j}$) yields

$$\det\left(\left(a_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) = \sum_{\sigma\in S_n}(-1)^\sigma \prod_{i=1}^n a_i^{g_{\sigma(i)}} = \sum_{\sigma\in S_n}\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)(-1)^\sigma .$$

This proves (1454).

[654] *Proof of (1455):* Let $\sigma \in S_n$. Then,

$$\left(\underbrace{\sum_{i=1}^n a_i b_{\sigma(i)}}_{=a_1b_{\sigma(1)}+a_2b_{\sigma(2)}+\cdots+a_nb_{\sigma(n)}}\right)^k = \left(a_1b_{\sigma(1)} + a_2b_{\sigma(2)} + \cdots + a_nb_{\sigma(n)}\right)^k$$
$$= \sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}} \mathbf{m}(k_1, k_2, \ldots, k_n) \prod_{i=1}^n \left(a_i b_{\sigma(i)}\right)^{k_i}$$

(by Exercise 6.2 (applied to n, k and $a_i b_{\sigma(i)}$ instead of m, n and a_i)). This proves (1455).

Hence,

$$\begin{aligned}
&\sum_{\sigma\in S_n}(-1)^\sigma \underbrace{\left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^k}_{\substack{=\sum\limits_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}} \mathbf{m}(k_1,k_2,\ldots,k_n)\prod\limits_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}\\ \text{(by (1455))}}}\\
&=\sum_{\sigma\in S_n}(-1)^\sigma \sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}} \mathbf{m}(k_1,k_2,\ldots,k_n)\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}\\
&=\underbrace{\sum_{\sigma\in S_n}\sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}}}_{=\sum\limits_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}}\sum\limits_{\sigma\in S_n}} (-1)^\sigma\,\mathbf{m}(k_1,k_2,\ldots,k_n)\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}\\
&=\sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}}\sum_{\sigma\in S_n}(-1)^\sigma\,\mathbf{m}(k_1,k_2,\ldots,k_n)\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}\\
&=\sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k}}\mathbf{m}(k_1,k_2,\ldots,k_n)\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}\\
&=\underbrace{\sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k;\\ (k_1,k_2,\ldots,k_n)\in\mathbf{E}}}}_{\substack{=\sum\limits_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ (k_1,k_2,\ldots,k_n)\in\mathbf{E};\\ k_1+k_2+\cdots+k_n=k}}=\sum\limits_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbf{E};\\ k_1+k_2+\cdots+k_n=k}}\\ \text{(since }\mathbf{E}\subseteq\mathbb{N}^n)}}\mathbf{m}(k_1,k_2,\ldots,k_n)\underbrace{\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}}_{\substack{=\left(\prod\limits_{i=1}^n a_i^{k_i}\right)\det\left(\left(b_i^{k_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)\\ \text{(by Lemma 7.239 }\textbf{(a)}\text{)}}}\\
&\qquad+\sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k;\\ (k_1,k_2,\ldots,k_n)\notin\mathbf{E}}}\mathbf{m}(k_1,k_2,\ldots,k_n)\underbrace{\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^n\left(a_ib_{\sigma(i)}\right)^{k_i}}_{\substack{=0\\ \text{(by (1452))}}}\\
&\qquad\left(\begin{array}{c}\text{since each }(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n\text{ satisfies either }(k_1,k_2,\ldots,k_n)\in\mathbf{E}\\ \text{or }(k_1,k_2,\ldots,k_n)\notin\mathbf{E}\text{ (but not both)}\end{array}\right)
\end{aligned}$$

$$= \sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbf{E};\\ k_1+k_2+\cdots+k_n=k}} \mathbf{m}(k_1,k_2,\ldots,k_n) \left(\prod_{i=1}^n a_i^{k_i}\right) \det\left(\left(b_i^{k_j}\right)_{1\le i\le n,\ 1\le j\le n}\right) + \underbrace{\sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbb{N}^n;\\ k_1+k_2+\cdots+k_n=k;\\ (k_1,k_2,\ldots,k_n)\notin\mathbf{E}}} \mathbf{m}(k_1,k_2,\ldots,k_n)\, 0}_{=0}$$

$$= \sum_{\substack{(k_1,k_2,\ldots,k_n)\in\mathbf{E};\\ k_1+k_2+\cdots+k_n=k}} \mathbf{m}(k_1,k_2,\ldots,k_n) \left(\prod_{i=1}^n a_i^{k_i}\right) \det\left(\left(b_i^{k_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)$$

$$= \sum_{\substack{((g_1,g_2,\ldots,g_n),\sigma)\in\mathbf{I}\times S_n;\\ g_{\sigma(1)}+g_{\sigma(2)}+\cdots+g_{\sigma(n)}=k}} \mathbf{m}\left(g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}\right) \left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right) \det\left(\left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}\right) \tag{1456}$$

(here, we have substituted $\left(g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}\right)$ for $(k_1,k_2,\ldots,k_n)$ in the sum, because the map

$$\mathbf{I}\times S_n \to \mathbf{E},$$
$$((g_1,g_2,\ldots,g_n),\sigma) \mapsto \left(g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}\right)$$

is a bijection).

But we have the following equality of summation signs:

$$\sum_{\substack{((g_1,g_2,\ldots,g_n),\sigma)\in\mathbf{I}\times S_n;\\ g_{\sigma(1)}+g_{\sigma(2)}+\cdots+g_{\sigma(n)}=k}} = \underbrace{\sum_{(g_1,g_2,\ldots,g_n)\in\mathbf{I}}}_{\substack{=\sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n}}\\ \text{(since }\mathbf{I}=\{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n\ \mid\ g_1<g_2<\cdots<g_n\}\text{)}}} \underbrace{\sum_{\substack{\sigma\in S_n;\\ g_{\sigma(1)}+g_{\sigma(2)}+\cdots+g_{\sigma(n)}=k}}}_{\substack{=\sum_{\substack{\sigma\in S_n;\\ g_1+g_2+\cdots+g_n=k}}\\ \text{(since each }\sigma\in S_n\text{ satisfies}\\ g_{\sigma(1)}+g_{\sigma(2)}+\cdots+g_{\sigma(n)}=g_1+g_2+\cdots+g_n\\ \text{(by Lemma 7.240 (a)))}}}$$

$$= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n}} \sum_{\substack{\sigma\in S_n;\\ g_1+g_2+\cdots+g_n=k}} = \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}} \sum_{\sigma\in S_n}.$$

Hence, (1456) becomes

$$\begin{aligned}
&\sum_{\sigma\in S_n}(-1)^\sigma\left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^k \\
&= \underbrace{\sum_{\substack{((g_1,g_2,\ldots,g_n),\sigma)\in\mathbf{I}\times S_n;\\ g_{\sigma(1)}+g_{\sigma(2)}+\cdots+g_{\sigma(n)}=k}}}_{=\sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}}\sum_{\sigma\in S_n}} \mathbf{m}\left(g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}\right)\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)\det\left(\left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}\right) \\
&= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}}\sum_{\sigma\in S_n}\underbrace{\mathbf{m}\left(g_{\sigma(1)},g_{\sigma(2)},\ldots,g_{\sigma(n)}\right)}_{\substack{=\mathbf{m}(g_1,g_2,\ldots,g_n)\\ \text{(by Lemma 7.240 (b))}}}\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)\underbrace{\det\left(\left(b_i^{g_{\sigma(j)}}\right)_{1\le i\le n,\ 1\le j\le n}\right)}_{\substack{=(-1)^\sigma\cdot\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)\\ \text{(by (1453))}}} \\
&= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}}\underbrace{\sum_{\sigma\in S_n}\mathbf{m}\left(g_1,g_2,\ldots,g_n\right)\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)(-1)^\sigma\cdot\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)}_{=\mathbf{m}(g_1,g_2,\ldots,g_n)\cdot\left(\sum_{\sigma\in S_n}\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)(-1)^\sigma\right)\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)} \\
&= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}}\mathbf{m}\left(g_1,g_2,\ldots,g_n\right)\cdot\underbrace{\left(\sum_{\sigma\in S_n}\left(\prod_{i=1}^n a_i^{g_{\sigma(i)}}\right)(-1)^\sigma\right)}_{\substack{=\det\left(\left(a_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)\\ \text{(by (1454))}}}\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right) \\
&= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}}\mathbf{m}\left(g_1,g_2,\ldots,g_n\right)\cdot\det\left(\left(a_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)\cdot\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right).
\end{aligned}$$

This proves Lemma 7.241. □

Next, we can use Lemma 7.241 to solve part **(a)** of Exercise 6.55:

Lemma 7.242. Let $n\in\mathbb{N}$. Let $a_1,a_2,\ldots,a_n$ be n elements of $\mathbb{K}$. Let $b_1,b_2,\ldots,b_n$ be n elements of $\mathbb{K}$. Let $m=\binom{n}{2}$.

Then,

$$\sum_{\sigma\in S_n}(-1)^\sigma\left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^k = 0$$

for each $k\in\{0,1,\ldots,m-1\}$.

Proof of Lemma 7.242. Let $k \in \{0, 1, \ldots, m-1\}$. Then, $k \in \mathbb{N}$ and $k \leq m-1$.

There exists no $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = k$ [655]. Thus, the sum

$$\sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}} \mathbf{m}(g_1, g_2, \ldots, g_n) \cdot \det\left(\left(a_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \cdot \det\left(\left(b_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$

is an empty sum. Hence,

$$\sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}} \mathbf{m}(g_1, g_2, \ldots, g_n) \cdot \det\left(\left(a_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \cdot \det\left(\left(b_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$
$$= (\text{empty sum}) = 0.$$

Now, Lemma 7.241 yields

$$\sum_{\sigma\in S_n} (-1)^\sigma \left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^k$$
$$= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=k}} \mathbf{m}(g_1, g_2, \ldots, g_n) \cdot \det\left(\left(a_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right) \cdot \det\left(\left(b_i^{g_j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}\right)$$
$$= 0.$$

This proves Lemma 7.242. □

Finally, we are ready for part **(b)** of Exercise 6.55:

Lemma 7.243. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$.

Let $m = \dbinom{n}{2}$. Then,

$$\sum_{\sigma\in S_n} (-1)^\sigma \left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^m = \mathbf{m}(0, 1, \ldots, n-1) \cdot \prod_{1\leq i<j\leq n} \left(\left(a_i - a_j\right)\left(b_i - b_j\right)\right).$$

Here, we are using the notation introduced in Exercise 6.2.

[655] *Proof.* Let $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ be such that $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = k$. We shall derive a contradiction.

Lemma 7.238 **(a)** (applied to $k_i = g_i$) yields $g_1 + g_2 + \cdots + g_n \geq \dbinom{n}{2}$. This contradicts $g_1 + g_2 + \cdots + g_n = k \leq m - 1 < m = \dbinom{n}{2}$.

Now, let us forget that we fixed $(g_1, g_2, \ldots, g_n)$. We thus have derived a contradiction for each $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = k$. Hence, there exists no $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = k$.

Proof of Lemma 7.243. We have $m = \dbinom{n}{2} = \sum_{r=0}^{n-1} r$ (by Lemma 7.184), thus $m = \sum_{r=0}^{n-1} r \in \mathbb{N}$. Hence, Lemma 7.241 (applied to $k = m$) yields

$$\sum_{\sigma \in S_n} (-1)^\sigma \left(\sum_{i=1}^n a_i b_{\sigma(i)} \right)^m = \sum_{\substack{(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n; \\ g_1 < g_2 < \cdots < g_n; \\ g_1 + g_2 + \cdots + g_n = m}} \mathbf{m}(g_1, g_2, \ldots, g_n) \cdot \det \left(\left(a_i^{g_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) \cdot \det \left(\left(b_i^{g_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right).$$

But we have the following two observations:

- The n-tuple $(0, 1, \ldots, n-1)$ is an n-tuple $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = m$ [656].
- Any n-tuple $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = m$ must be equal to $(0, 1, \ldots, n-1)$ [657].

Combining these two observations, we conclude that there exists exactly one n-tuple $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = m$: namely, the n-tuple $(0, 1, \ldots, n-1)$. Hence, the sum

$$\sum_{\substack{(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n; \\ g_1 < g_2 < \cdots < g_n; \\ g_1 + g_2 + \cdots + g_n = m}} \mathbf{m}(g_1, g_2, \ldots, g_n) \cdot \det \left(\left(a_i^{g_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right) \cdot \det \left(\left(b_i^{g_j} \right)_{1 \leq i \leq n,\ 1 \leq j \leq n} \right)$$

has only one addend: namely, the addend for $(g_1, g_2, \ldots, g_n) = (0, 1, \ldots, n-1)$.

[656] *Proof.* The n-tuple $(0, 1, \ldots, n-1)$ belongs to $\mathbb{N}^n$ and satisfies $0 < 1 < \cdots < n-1$ and $0 + 1 + \cdots + (n-1) = m$ (since $0 + 1 + \cdots + (n-1) = \sum_{r=0}^{n-1} r = m$). In other words, the n-tuple $(0, 1, \ldots, n-1)$ is an n-tuple $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = m$.

[657] *Proof.* Let $(g_1, g_2, \ldots, g_n) \in \mathbb{N}^n$ be any n-tuple satisfying $g_1 < g_2 < \cdots < g_n$ and $g_1 + g_2 + \cdots + g_n = m$. We shall show that $(g_1, g_2, \ldots, g_n)$ must be equal to $(0, 1, \ldots, n-1)$.

We have $g_1 + g_2 + \cdots + g_n = m \leq m = \dbinom{n}{2}$. Hence, Lemma 7.238 **(b)** (applied to g_i instead of k_i) shows that $(g_1, g_2, \ldots, g_n) = (0, 1, \ldots, n-1)$. In other words, $(g_1, g_2, \ldots, g_n)$ must be equal to $(0, 1, \ldots, n-1)$. Qed.

Therefore, this sum can be simplified as follows:

$$\sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=m}} \mathbf{m}(g_1,g_2,\ldots,g_n)\cdot\det\left(\left(a_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)\cdot\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)$$

$$= \mathbf{m}(0,1,\ldots,n-1)\cdot\underbrace{\det\left(\left(a_i^{j-1}\right)_{1\le i\le n,\ 1\le j\le n}\right)}_{\substack{=\prod_{1\le j<i\le n}(a_i-a_j)\\ \text{(by Theorem 6.46 (c)}\\ \text{(applied to } x_k=a_k))}}\cdot\underbrace{\det\left(\left(b_i^{j-1}\right)_{1\le i\le n,\ 1\le j\le n}\right)}_{\substack{=\prod_{1\le j<i\le n}(b_i-b_j)\\ \text{(by Theorem 6.46 (c)}\\ \text{(applied to } x_k=b_k))}}$$

$$= \mathbf{m}(0,1,\ldots,n-1)\cdot\underbrace{\left(\prod_{1\le j<i\le n}(a_i-a_j)\right)\cdot\left(\prod_{1\le j<i\le n}(b_i-b_j)\right)}_{\substack{=\prod_{1\le j<i\le n}((a_i-a_j)(b_i-b_j))\\ =\prod_{1\le i<j\le n}((a_j-a_i)(b_j-b_i))\\ \text{(here, we have renamed the index } (j,i) \text{ as } (i,j)\\ \text{in the product)}}}$$

$$= \mathbf{m}(0,1,\ldots,n-1)\cdot\prod_{1\le i<j\le n}\left(\underbrace{(a_j-a_i)(b_j-b_i)}_{=(a_i-a_j)(b_i-b_j)}\right)$$

$$= \mathbf{m}(0,1,\ldots,n-1)\cdot\prod_{1\le i<j\le n}\left((a_i-a_j)(b_i-b_j)\right).$$

Thus,

$$\sum_{\sigma\in S_n}(-1)^\sigma\left(\sum_{i=1}^n a_i b_{\sigma(i)}\right)^m$$

$$= \sum_{\substack{(g_1,g_2,\ldots,g_n)\in\mathbb{N}^n;\\ g_1<g_2<\cdots<g_n;\\ g_1+g_2+\cdots+g_n=m}} \mathbf{m}(g_1,g_2,\ldots,g_n)\cdot\det\left(\left(a_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)\cdot\det\left(\left(b_i^{g_j}\right)_{1\le i\le n,\ 1\le j\le n}\right)$$

$$= \mathbf{m}(0,1,\ldots,n-1)\cdot\prod_{1\le i<j\le n}\left((a_i-a_j)(b_i-b_j)\right).$$

This proves Lemma 7.243. □

Solution to Exercise 6.55. Exercise 6.55 **(a)** follows from Lemma 7.242. Exercise 6.55 **(b)** follows from Lemma 7.243. Exercise 6.55 **(c)** follows from Lemma 7.241. □

7.123.2. Additional observations

Let us also record two corollaries of Lemma 7.238:

Corollary 7.244. Let $n \in \mathbb{N}$. Let $(p_1, p_2, \ldots, p_n) \in \mathbb{N}^n$ be such that $p_1, p_2, \ldots, p_n$ are distinct. Then, $p_1 + p_2 + \cdots + p_n \geq \binom{n}{2}$.

Proof of Corollary 7.244. Define the subsets $\mathbf{E}$ and $\mathbf{I}$ of $\mathbb{N}^n$ as in Lemma 7.237. Then, Lemma 7.237 shows that the map

$$\mathbf{I} \times S_n \to \mathbf{E},$$
$$((g_1, g_2, \ldots, g_n), \sigma) \mapsto \left(g_{\sigma(1)}, g_{\sigma(2)}, \ldots, g_{\sigma(n)}\right)$$

is well-defined and is a bijection. Denote this map by Φ.

Notice that

$$\mathbf{E} = \{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct}\} \tag{1457}$$

(by the definition of $\mathbf{E}$).

Now, $(p_1, p_2, \ldots, p_n)$ is an element of $\mathbb{N}^n$ having the property that the integers $p_1, p_2, \ldots, p_n$ are distinct. In other words,

$$(p_1, p_2, \ldots, p_n) \in \{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n \mid \text{the integers } k_1, k_2, \ldots, k_n \text{ are distinct}\}.$$

In view of (1457), this rewrites as $(p_1, p_2, \ldots, p_n) \in \mathbf{E}$.

But the map Φ is surjective (since Φ is a bijection). In other words, $\mathbf{E} = \Phi(\mathbf{I} \times S_n)$. Thus, $(p_1, p_2, \ldots, p_n) \in \mathbf{E} = \Phi(\mathbf{I} \times S_n)$. In other words, there exists some $((k_1, k_2, \ldots, k_n), \sigma) \in \mathbf{I} \times S_n$ such that $(p_1, p_2, \ldots, p_n) = \Phi(((k_1, k_2, \ldots, k_n), \sigma))$. Consider this $((k_1, k_2, \ldots, k_n), \sigma)$. We have $(k_1, k_2, \ldots, k_n) \in \mathbf{I}$, and thus $k_1 < k_2 < \cdots < k_n$ (by the definition of $\mathbf{I}$).

We have

$$(p_1, p_2, \ldots, p_n) = \Phi(((k_1, k_2, \ldots, k_n), \sigma)) = \left(k_{\sigma(1)}, k_{\sigma(2)}, \ldots, k_{\sigma(n)}\right)$$

(by the definition of the map Φ). Thus,

$$\begin{aligned} p_1 + p_2 + \cdots + p_n &= k_{\sigma(1)} + k_{\sigma(2)} + \cdots + k_{\sigma(n)} = k_1 + k_2 + \cdots + k_n \\ &\qquad \left(\text{by Lemma 7.240 } \textbf{(a)}, \text{ applied to } g_i = k_i\right) \\ &\geq \binom{n}{2} \qquad \left(\text{by Lemma 7.238 } \textbf{(a)}\right). \end{aligned}$$

This proves Corollary 7.244. □

Corollary 7.245. Let $n \in \mathbb{N}$. Let $a_1, a_2, \ldots, a_n$ be n elements of $\mathbb{K}$. Let $b_1, b_2, \ldots, b_n$ be n elements of $\mathbb{K}$.

Let $(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$ be such that $k_1 + k_2 + \cdots + k_n < \binom{n}{2}$. Then,

$$\sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} \left(a_i b_{\sigma(i)}\right)^{k_i} = 0.$$

Proof of Corollary 7.245. The integers $k_1, k_2, \ldots, k_n$ are not distinct[658]. Hence, Lemma 7.239 **(b)** yields $\sum\limits_{\sigma \in S_n} (-1)^{\sigma} \prod\limits_{i=1}^{n} \left(a_i b_{\sigma(i)}\right)^{k_i} = 0$. This proves Corollary 7.245. □

7.124. Solution to Exercise 6.56

7.124.1. First solution

Our first solution to Exercise 6.56 (inspired by [BruRys91, Lemma 9.2.10]) shall follow the hint given. We are going to prepare for it by stating several simple lemmas.

First, let us agree on some notations. We are going to use the notations introduced in Definition 6.78, in Definition 6.153 and in Definition 7.174 throughout Section 7.124.

Now, we collect some useful lemmas.

Lemma 7.246. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times m}$. Let $\gamma \in S_n$ and $\delta \in S_m$. Let $i_1, i_2, \ldots, i_u$ be some elements of $\{1, 2, \ldots, n\}$; let $j_1, j_2, \ldots, j_v$ be some elements of $\{1, 2, \ldots, m\}$. Then,

$$\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} \left(A_{[\gamma, \delta]}\right) = \operatorname{sub}_{\gamma(i_1), \gamma(i_2), \ldots, \gamma(i_u)}^{\delta(j_1), \delta(j_2), \ldots, \delta(j_v)} A.$$

Proof of Lemma 7.246. Write the $n \times m$-matrix A in the form $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Then,

$$\operatorname{sub}_{\gamma(i_1), \gamma(i_2), \ldots, \gamma(i_u)}^{\delta(j_1), \delta(j_2), \ldots, \delta(j_v)} A = \left(a_{\gamma(i_x), \delta(j_y)}\right)_{1 \leq x \leq u,\ 1 \leq y \leq v} \tag{1458}$$

(by the definition of $\operatorname{sub}_{\gamma(i_1), \gamma(i_2), \ldots, \gamma(i_u)}^{\delta(j_1), \delta(j_2), \ldots, \delta(j_v)} A$, since $A = \left(a_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$). On the other hand, the definition of $A_{[\gamma, \delta]}$ yields $A_{[\gamma, \delta]} = \left(a_{\gamma(i), \delta(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Hence, the definition of $\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} \left(A_{[\gamma, \delta]}\right)$ yields

$$\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} \left(A_{[\gamma, \delta]}\right) = \left(a_{\gamma(i_x), \delta(j_y)}\right)_{1 \leq x \leq u,\ 1 \leq y \leq v}.$$

Comparing this with (1458), we obtain $\operatorname{sub}_{i_1, i_2, \ldots, i_u}^{j_1, j_2, \ldots, j_v} \left(A_{[\gamma, \delta]}\right) = \operatorname{sub}_{\gamma(i_1), \gamma(i_2), \ldots, \gamma(i_u)}^{\delta(j_1), \delta(j_2), \ldots, \delta(j_v)} A$. This proves Lemma 7.246. □

[658] *Proof.* Assume the contrary. Thus, the integers $k_1, k_2, \ldots, k_n$ are distinct. Hence, Corollary 7.244 (applied to k_i instead of p_i) shows that $k_1 + k_2 + \cdots + k_n \geq \dbinom{n}{2}$. This contradicts $k_1 + k_2 + \cdots + k_n < \dbinom{n}{2}$. This contradiction shows that our assumption was wrong. Qed.

Lemma 7.247. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let $\gamma \in S_n$, $\delta \in S_m$ and $\varepsilon \in S_p$. Let $A \in \mathbb{K}^{n \times m}$ and $B \in \mathbb{K}^{m \times p}$. Then,

$$(AB)_{[\gamma,\varepsilon]} = A_{[\gamma,\delta]} B_{[\delta,\varepsilon]}.$$

Proof of Lemma 7.247. Write the $n \times m$-matrix A in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m}$. Then, $A_{[\gamma,\delta]} = \left(a_{\gamma(i),\delta(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$ (by the definition of $A_{[\gamma,\delta]}$).

Write the $m \times p$-matrix B in the form $B = (b_{i,j})_{1 \leq i \leq m,\ 1 \leq j \leq p}$. Then, $B_{[\delta,\varepsilon]} = \left(b_{\delta(i),\varepsilon(j)}\right)_{1 \leq i \leq m,\ 1 \leq j \leq p}$ (by the definition of $B_{[\delta,\varepsilon]}$).

Every $(i,j) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,p\}$ satisfies

$$\sum_{k=1}^{m} a_{\gamma(i),\delta(k)} b_{\delta(k),\varepsilon(j)} = \sum_{k=1}^{m} a_{\gamma(i),k} b_{k,\varepsilon(j)}$$

(here, we have substituted k for $\delta(k)$ in the sum, since the map $\delta : \{1,2,\ldots,m\} \to \{1,2,\ldots,m\}$ is a bijection). In other words, we have

$$\left(\sum_{k=1}^{m} a_{\gamma(i),\delta(k)} b_{\delta(k),\varepsilon(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p} = \left(\sum_{k=1}^{m} a_{\gamma(i),k} b_{k,\varepsilon(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p}. \tag{1459}$$

The definition of the product of two matrices yields

$$\begin{aligned} A_{[\gamma,\delta]} B_{[\delta,\varepsilon]} &= \left(\sum_{k=1}^{m} a_{\gamma(i),\delta(k)} b_{\delta(k),\varepsilon(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p} \\ &\qquad \left(\begin{array}{c} \text{since } A_{[\gamma,\delta]} = \left(a_{\gamma(i),\delta(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m} \\ \text{and } B_{[\delta,\varepsilon]} = \left(b_{\delta(i),\varepsilon(j)}\right)_{1 \leq i \leq m,\ 1 \leq j \leq p} \end{array}\right) \\ &= \left(\sum_{k=1}^{m} a_{\gamma(i),k} b_{k,\varepsilon(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p} \qquad (\text{by (1459)}). \end{aligned} \tag{1460}$$

On the other hand, the definition of the product of two matrices yields $AB = \left(\sum_{k=1}^{m} a_{i,k} b_{k,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p}$ (since $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m}$ and $B = (b_{i,j})_{1 \leq i \leq m,\ 1 \leq j \leq p}$). Hence, the definition of $(AB)_{[\gamma,\varepsilon]}$ yields

$$(AB)_{[\gamma,\varepsilon]} = \left(\sum_{k=1}^{m} a_{\gamma(i),k} b_{k,\varepsilon(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p}.$$

Comparing this with (1460), we obtain $(AB)_{[\gamma,\varepsilon]} = A_{[\gamma,\delta]} B_{[\delta,\varepsilon]}$. Thus, Lemma 7.247 is proven. $\square$

Lemma 7.248. Let $n \in \mathbb{N}$ and $\gamma \in S_n$. Then, $(I_n)_{[\gamma,\gamma]} = I_n$.

Proof of Lemma 7.248. For every two objects i and j, define an element $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$. Then, $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of I_n). Hence, $(I_n)_{[\gamma,\gamma]} = \left(\delta_{\gamma(i),\gamma(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of $(I_n)_{[\gamma,\gamma]}$).

The map γ is a permutation (since $\gamma \in S_n$), thus bijective, thus injective.

For every $(i,j) \in \{1,2,\ldots,n\}^2$, we have $\delta_{\gamma(i),\gamma(j)} = \delta_{i,j}$ [659]. In other words, $\left(\delta_{\gamma(i),\gamma(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence,

$$(I_n)_{[\gamma,\gamma]} = \left(\delta_{\gamma(i),\gamma(j)}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

Comparing this with $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, we obtain $(I_n)_{[\gamma,\gamma]} = I_n$. This proves Lemma 7.248. □

Lemma 7.249. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times n}$ be an invertible matrix. Let $\gamma \in S_n$ and $\delta \in S_n$. Then, the matrix $A_{[\gamma,\delta]} \in \mathbb{K}^{n \times n}$ is invertible, and its inverse is $\left(A_{[\gamma,\delta]}\right)^{-1} = \left(A^{-1}\right)_{[\delta,\gamma]}$.

Proof of Lemma 7.249. Lemma 7.247 (applied to n, n, γ and A^{-1} instead of m, p, ε and B) yields $\left(AA^{-1}\right)_{[\gamma,\gamma]} = A_{[\gamma,\delta]} \left(A^{-1}\right)_{[\delta,\gamma]}$. Thus,

$$A_{[\gamma,\delta]} \left(A^{-1}\right)_{[\delta,\gamma]} = \left(\underbrace{AA^{-1}}_{=I_n}\right)_{[\gamma,\gamma]} = (I_n)_{[\gamma,\gamma]} = I_n \qquad (1461)$$

(by Lemma 7.248). Also, Lemma 7.247 (applied to n, n, δ, γ, δ, A^{-1} and A instead of m, p, γ, δ, ε, A and B) yields $\left(A^{-1}A\right)_{[\delta,\delta]} = \left(A^{-1}\right)_{[\delta,\gamma]} A_{[\gamma,\delta]}$. Thus,

$$\left(A^{-1}\right)_{[\delta,\gamma]} A_{[\gamma,\delta]} = \left(\underbrace{A^{-1}A}_{=I_n}\right)_{[\delta,\delta]} = (I_n)_{[\delta,\delta]} = I_n \qquad (1462)$$

[659] *Proof.* Let $(i,j) \in \{1,2,\ldots,n\}^2$. We must prove $\delta_{\gamma(i),\gamma(j)} = \delta_{i,j}$.

If $i = j$, then both $\delta_{\gamma(i),\gamma(j)}$ and $\delta_{i,j}$ equal 1 (because $\gamma\left(\underbrace{i}_{=j}\right) = \gamma(j)$ shows that $\delta_{\gamma(i),\gamma(j)} = 1$, whereas $i = j$ shows that $\delta_{i,j} = 1$). Hence, if $i = j$, then $\delta_{\gamma(i),\gamma(j)} = \delta_{i,j}$ holds. Thus, for the rest of our proof of $\delta_{\gamma(i),\gamma(j)} = \delta_{i,j}$, we WLOG assume that $i \neq j$. Hence, $\delta_{i,j} = 0$.

But the map γ is injective. Thus, from $i \neq j$, we obtain $\gamma(i) \neq \gamma(j)$. Hence, $\delta_{\gamma(i),\gamma(j)} = 0$. Comparing this with $\delta_{i,j} = 0$, we obtain $\delta_{\gamma(i),\gamma(j)} = \delta_{i,j}$, qed.

(by Lemma 7.248 (applied to δ instead of γ)).

The two equalities (1461) and (1462) (combined) show that the matrix $\left(A^{-1}\right)_{[\delta,\gamma]}$ is an inverse of the matrix $A_{[\gamma,\delta]}$. Thus, the matrix $A_{[\gamma,\delta]} \in \mathbb{K}^{n\times n}$ is invertible, and its inverse is $\left(A_{[\gamma,\delta]}\right)^{-1} = \left(A^{-1}\right)_{[\delta,\gamma]}$. Lemma 7.249 is proven. □

We can finally step to the solution of Exercise 6.56:

First solution to Exercise 6.56. Let $k = |P|$. Thus, $k = |P| = |Q|$.

Clearly, $k \in \{0, 1, \ldots, n\}$ (since $k = |P|$ for a subset P of $\{1, 2, \ldots, n\}$).

Lemma 7.214 (applied to $I = P$) yields that there exists a $\sigma \in S_n$ satisfying $(\sigma(1), \sigma(2), \ldots, \sigma(k)) = w(P)$, $(\sigma(k+1), \sigma(k+2), \ldots, \sigma(n)) = w\left(\widetilde{P}\right)$ and $(-1)^{\sigma} = (-1)^{\sum P - (1+2+\cdots+k)}$. Denote this σ by γ. Thus, γ is an element of S_n satisfying $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(P)$, $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{P}\right)$ and $(-1)^{\gamma} = (-1)^{\sum P - (1+2+\cdots+k)}$.

Lemma 7.214 (applied to $I = Q$) yields that there exists a $\sigma \in S_n$ satisfying $(\sigma(1), \sigma(2), \ldots, \sigma(k)) = w(Q)$, $(\sigma(k+1), \sigma(k+2), \ldots, \sigma(n)) = w\left(\widetilde{Q}\right)$ and $(-1)^{\sigma} = (-1)^{\sum Q - (1+2+\cdots+k)}$. Denote this σ by δ. Thus, δ is an element of S_n satisfying $(\delta(1), \delta(2), \ldots, \delta(k)) = w(Q)$, $(\delta(k+1), \delta(k+2), \ldots, \delta(n)) = w\left(\widetilde{Q}\right)$ and $(-1)^{\delta} = (-1)^{\sum Q - (1+2+\cdots+k)}$.

Lemma 7.249 shows that the matrix $A_{[\gamma,\delta]} \in \mathbb{K}^{n\times n}$ is invertible, and its inverse is $\left(A_{[\gamma,\delta]}\right)^{-1} = \left(A^{-1}\right)_{[\delta,\gamma]}$. Hence, Exercise 6.38 (applied to $A_{[\gamma,\delta]}$ instead of A) yields that

$$\begin{aligned}
&\det\left(\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A_{[\gamma,\delta]}\right)\right) \\
&= \det\left(A_{[\gamma,\delta]}\right) \cdot \det\left(\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(\underbrace{\left(A_{[\gamma,\delta]}\right)^{-1}}_{=\left(A^{-1}\right)_{[\delta,\gamma]}}\right)\right) \\
&= \det\left(A_{[\gamma,\delta]}\right) \cdot \det\left(\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(\left(A^{-1}\right)_{[\delta,\gamma]}\right)\right). \qquad (1463)
\end{aligned}$$

But Lemma 7.246 (applied to n, k, k, $(1, 2, \ldots, k)$ and $(1, 2, \ldots, k)$ instead of m, u, v, $(i_1, i_2, \ldots, i_u)$ and $(j_1, j_2, \ldots, j_v)$) yields

$$\begin{aligned}
\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A_{[\gamma,\delta]}\right) &= \operatorname{sub}_{\gamma(1),\gamma(2),\ldots,\gamma(k)}^{\delta(1),\delta(2),\ldots,\delta(k)} A = \operatorname{sub}_{(\gamma(1),\gamma(2),\ldots,\gamma(k))}^{(\delta(1),\delta(2),\ldots,\delta(k))} A \\
&= \operatorname{sub}_{w(P)}^{w(Q)} A \qquad (1464)
\end{aligned}$$

(since $(\gamma(1), \gamma(2), \ldots, \gamma(k)) = w(P)$ and $(\delta(1), \delta(2), \ldots, \delta(k)) = w(Q)$).

Furthermore, we have $A^{-1} \in \mathbb{K}^{n\times n}$ (since $A \in \mathbb{K}^{n\times n}$). Hence, Lemma 7.246 (applied to n, A^{-1}, δ, γ, $n-k$, $n-k$, $(k+1, k+2, \ldots, n)$ and $(k+1, k+2, \ldots, n)$ instead of m, A, γ, δ, u, v, $(i_1, i_2, \ldots, i_u)$ and $(j_1, j_2, \ldots, j_v)$) yields

$$\begin{aligned} \operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(\left(A^{-1}\right)_{[\delta,\gamma]}\right) &= \operatorname{sub}_{\delta(k+1),\delta(k+2),\ldots,\delta(n)}^{\gamma(k+1),\gamma(k+2),\ldots,\gamma(n)}\left(A^{-1}\right) \\ &= \operatorname{sub}_{(\delta(k+1),\delta(k+2),\ldots,\delta(n))}^{(\gamma(k+1),\gamma(k+2),\ldots,\gamma(n))}\left(A^{-1}\right) \\ &= \operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})}\left(A^{-1}\right) \end{aligned} \qquad (1465)$$

(since $(\delta(k+1), \delta(k+2), \ldots, \delta(n)) = w\left(\widetilde{Q}\right)$ and $(\gamma(k+1), \gamma(k+2), \ldots, \gamma(n)) = w\left(\widetilde{P}\right)$).

On the other hand,

$$\begin{aligned} &\underbrace{(-1)^{\gamma}}_{=(-1)^{\sum P-(1+2+\cdots+k)}} \underbrace{(-1)^{\delta}}_{=(-1)^{\sum Q-(1+2+\cdots+k)}} \\ &= (-1)^{\sum P-(1+2+\cdots+k)} (-1)^{\sum Q-(1+2+\cdots+k)} = (-1)^{(\sum P-(1+2+\cdots+k))+(\sum Q-(1+2+\cdots+k))} \\ &= (-1)^{\sum P+\sum Q} \end{aligned} \qquad (1466)$$

(since

$$\begin{aligned} &\left(\sum P-(1+2+\cdots+k)\right)+\left(\sum Q-(1+2+\cdots+k)\right) \\ &= \left(\sum P+\sum Q\right)-2(1+2+\cdots+k) \equiv \sum P+\sum Q \mod 2 \end{aligned}$$

).

Lemma 7.178 yields

$$\det\left(A_{[\gamma,\delta]}\right) = \underbrace{(-1)^{\gamma}(-1)^{\delta}}_{\substack{=(-1)^{\sum P+\sum Q} \\ \text{(by (1466))}}} \det A = (-1)^{\sum P+\sum Q} \det A. \qquad (1467)$$

Now,

$$\begin{aligned}
&\det\left(\underbrace{\operatorname{sub}_{w(P)}^{w(Q)} A}_{\substack{=\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A_{[\gamma,\delta]}\right) \\ \text{(by (1464))}}}\right) \\
&= \det\left(\operatorname{sub}_{1,2,\ldots,k}^{1,2,\ldots,k}\left(A_{[\gamma,\delta]}\right)\right) \\
&= \underbrace{\det\left(A_{[\gamma,\delta]}\right)}_{\substack{=(-1)^{\sum P+\sum Q}\det A \\ \text{(by (1467))}}} \cdot \det\left(\underbrace{\operatorname{sub}_{k+1,k+2,\ldots,n}^{k+1,k+2,\ldots,n}\left(\left(A^{-1}\right)_{[\delta,\gamma]}\right)}_{\substack{=\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})}\left(A^{-1}\right) \\ \text{(by (1465))}}}\right) \qquad \text{(by (1463))} \\
&= (-1)^{\sum P+\sum Q}\det A \cdot \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})}\left(A^{-1}\right)\right).
\end{aligned}$$

This solves Exercise 6.56. □

7.124.2. Second solution

Now, let us give a second solution to Exercise 6.56, following an idea from [LLPT95, Chapter SCHUR, proof of (1.9)].

Throughout this section, we shall use the notations introduced in Definition 6.78 and in Definition 6.153.

Let us first prove a simple combinatorial fact:

Lemma 7.250. Let S be a set of integers. Let $u \in \mathbb{N}$. Let $\mathcal{P}_u(S)$ denote the set of all u-element subsets of S. Let $\mathbf{I}$ denote the set

$$\left\{(g_1, g_2, \ldots, g_u) \in S^u \mid g_1 < g_2 < \cdots < g_u\right\}.$$

Then, the map

$$\mathcal{P}_u(S) \to \mathbf{I}, \qquad R \mapsto w(R)$$

is well-defined and a bijection.

Proof of Lemma 7.250. Recall that $\mathcal{P}_u(S)$ is the set of all u-element subsets of S, whereas $\mathbf{I}$ is the set of all strictly increasing lists[660] of u elements of S.

[660] A list $(g_1, g_2, \ldots, g_u)$ is said to be *strictly increasing* if it satisfies $g_1 < g_2 < \cdots < g_u$.

For every $R \in \mathcal{P}_u(S)$, the list $w(R)$ is the list of all elements of R in increasing order (with no repetitions). Thus, this list $w(R)$ is a strictly increasing list with $|R| = u$ entries (which all belong to R and therefore to S); hence, this list $w(R)$ belongs to $\mathbf{I}$. Thus, the map

$$\mathcal{P}_u(S) \to \mathbf{I}, \qquad R \mapsto w(R)$$

is well-defined. Also, the map

$$\mathbf{I} \to \mathcal{P}_u(S), \qquad (g_1, g_2, \ldots, g_u) \mapsto \{g_1, g_2, \ldots, g_u\}$$

is well-defined (because if $(g_1, g_2, \ldots, g_u) \in \mathbf{I}$, then we have $g_1 < g_2 < \cdots < g_u$, and thus the u elements $g_1, g_2, \ldots, g_u$ are pairwise distinct; therefore, $\{g_1, g_2, \ldots, g_u\}$ is a u-element set). These two maps are mutually inverse[661]. Hence, the map

$$\mathcal{P}_u(S) \to \mathbf{I}, \qquad R \mapsto w(R)$$

is invertible, i.e., is a bijection. □

We can use Lemma 7.250 to obtain the following (slightly weaker) form of Corollary 7.182:

Corollary 7.251. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let A be an $n \times p$-matrix. Let B be a $p \times m$-matrix. Let $k \in \mathbb{N}$. Let P be a subset of $\{1, 2, \ldots, n\}$ such that $|P| = k$. Let Q be a subset of $\{1, 2, \ldots, m\}$ such that $|Q| = k$. Then,

$$\det\left(\operatorname{sub}_{w(P)}^{w(Q)}(AB)\right) = \sum_{\substack{R \subseteq \{1,2,\ldots,p\};\\ |R|=k}} \det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right).$$

Proof of Corollary 7.251. For every set S, we let $\mathcal{P}_k(S)$ denote the set of all k-element subsets of S.

Recall that $w(P)$ was defined as the list of all elements of P in increasing order (with no repetitions). Thus, $w(P)$ is a list of $|P|$ elements. In other words, $w(P)$ is a list of k elements (since $|P| = k$). Similarly, $w(Q)$ is a list of k elements.

Write the list $w(P)$ in the form $w(P) = (i_1, i_2, \ldots, i_k)$. (This is possible, since $w(P)$ is a list of k elements.)

Write the list $w(Q)$ in the form $w(Q) = (j_1, j_2, \ldots, j_k)$. (This is possible, since $w(Q)$ is a list of k elements.)

[661] *Proof.* The former map takes a u-element subset R of S and lists its elements in increasing order; the latter map takes a strictly increasing list (of u elements of S) and outputs the set of its entries. These two operations clearly revert one another. Thus, the two maps are mutually inverse.

We have $(i_1, i_2, \ldots, i_k) = w(P)$. Thus, the elements $i_1, i_2, \ldots, i_k$ are elements of P, and hence also elements of $\{1, 2, \ldots, n\}$ (since $P \subseteq \{1, 2, \ldots, n\}$). The same argument (but applied to m, Q and $(j_1, j_2, \ldots, j_k)$ instead of n, P and $(i_1, i_2, \ldots, i_k)$) yields that $j_1, j_2, \ldots, j_k$ are elements of $\{1, 2, \ldots, m\}$. Hence, Corollary 7.182 (applied to $u = k$) yields

$$\det\left(\operatorname{sub}_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_k}(AB)\right) = \sum_{1 \leq g_1 < g_2 < \cdots < g_k \leq p} \det\left(\operatorname{sub}_{i_1,i_2,\ldots,i_k}^{g_1,g_2,\ldots,g_k} A\right) \cdot \det\left(\operatorname{sub}_{g_1,g_2,\ldots,g_k}^{j_1,j_2,\ldots,j_k} B\right). \tag{1468}$$

(Here, the summation sign "$\sum_{1 \leq g_1 < g_2 < \cdots < g_k \leq p}$" has to be interpreted as "$\sum_{\substack{(g_1,g_2,\ldots,g_k) \in \{1,2,\ldots,p\}^k; \\ g_1 < g_2 < \cdots < g_k}}$", in analogy to Remark 6.33.)

Now, let S denote the set $\{1, 2, \ldots, p\}$. Recall that $\mathcal{P}_k(S)$ denotes the set of all k-element subsets of S. In other words,

$$\mathcal{P}_k(S) = \{T \subseteq S \mid |T| = k\}.$$

Let $\mathbf{I}$ denote the set

$$\left\{(g_1, g_2, \ldots, g_k) \in S^k \mid g_1 < g_2 < \cdots < g_k\right\}.$$

Lemma 7.250 (applied to $u = k$) shows that the map

$$\mathcal{P}_k(S) \to \mathbf{I}, \qquad R \mapsto w(R)$$

is well-defined and a bijection.

The definition of $\mathbf{I}$ yields $\mathbf{I} = \left\{(g_1, g_2, \ldots, g_k) \in S^k \mid g_1 < g_2 < \cdots < g_k\right\}$. Hence, we have the following equality of summation signs:

$$\sum_{(g_1,g_2,\ldots,g_k) \in \mathbf{I}} = \sum_{\substack{(g_1,g_2,\ldots,g_k) \in S^k; \\ g_1 < g_2 < \cdots < g_k}} = \sum_{\substack{(g_1,g_2,\ldots,g_k) \in \{1,2,\ldots,p\}^k; \\ g_1 < g_2 < \cdots < g_k}}$$

(since $S = \{1, 2, \ldots, p\}$). Comparing this with

$$\sum_{1 \leq g_1 < g_2 < \cdots < g_k \leq p} = \sum_{\substack{(g_1,g_2,\ldots,g_k) \in \{1,2,\ldots,p\}^k; \\ g_1 < g_2 < \cdots < g_k}},$$

we obtain

$$\sum_{1 \leq g_1 < g_2 < \cdots < g_k \leq p} = \sum_{(g_1,g_2,\ldots,g_k) \in \mathbf{I}}.$$

Now, (1468) becomes

$$\det\left(\operatorname{sub}_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_k}(AB)\right)$$

$$= \underbrace{\sum_{1\leq g_1<g_2<\cdots<g_k\leq p}}_{=\sum_{(g_1,g_2,\ldots,g_k)\in\mathbf{I}}} \det\left(\underbrace{\operatorname{sub}_{i_1,i_2,\ldots,i_k}^{g_1,g_2,\ldots,g_k} A}_{=\operatorname{sub}_{(i_1,i_2,\ldots,i_k)}^{(g_1,g_2,\ldots,g_k)} A}\right)\cdot\det\left(\underbrace{\operatorname{sub}_{g_1,g_2,\ldots,g_k}^{j_1,j_2,\ldots,j_k} B}_{=\operatorname{sub}_{(g_1,g_2,\ldots,g_k)}^{(j_1,j_2,\ldots,j_k)} B}\right)$$

$$= \sum_{(g_1,g_2,\ldots,g_k)\in\mathbf{I}} \det\left(\operatorname{sub}_{(i_1,i_2,\ldots,i_k)}^{(g_1,g_2,\ldots,g_k)} A\right)\cdot\det\left(\operatorname{sub}_{(g_1,g_2,\ldots,g_k)}^{(j_1,j_2,\ldots,j_k)} B\right)$$

$$= \underbrace{\sum_{R\in\mathcal{P}_k(S)}}_{\substack{=\sum_{R\subseteq S;\ |R|=k} \\ (\text{since } \mathcal{P}_k(S)=\{T\subseteq S\ \mid\ |T|=k\})}} \det\left(\underbrace{\operatorname{sub}_{(i_1,i_2,\ldots,i_k)}^{w(R)} A}_{\substack{=\operatorname{sub}_{w(P)}^{w(R)} A \\ (\text{since } (i_1,i_2,\ldots,i_k)=w(P))}}\right)\cdot\det\left(\underbrace{\operatorname{sub}_{w(R)}^{(j_1,j_2,\ldots,j_k)} B}_{\substack{=\operatorname{sub}_{w(R)}^{w(Q)} B \\ (\text{since } (j_1,j_2,\ldots,j_k)=w(Q))}}\right)$$

$$\left(\begin{array}{c}\text{here, we have substituted } w(R) \text{ for } (g_1,g_2,\ldots,g_k)\text{, since} \\ \text{the map } \mathcal{P}_k(S)\to\mathbf{I},\ R\mapsto w(R) \text{ is a bijection}\end{array}\right)$$

$$= \underbrace{\sum_{R\subseteq S;\ |R|=k}}_{\substack{=\sum_{R\subseteq\{1,2,\ldots,p\};\ |R|=k} \\ (\text{since } S=\{1,2,\ldots,p\})}} \det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right)\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$$

$$= \sum_{\substack{R\subseteq\{1,2,\ldots,p\};\\ |R|=k}} \det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right)\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right).$$

Comparing this with

$$\det\left(\underbrace{\operatorname{sub}_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_k}(AB)}_{=\operatorname{sub}_{(i_1,i_2,\ldots,i_k)}^{(j_1,j_2,\ldots,j_k)}(AB)}\right) = \det\left(\operatorname{sub}_{(i_1,i_2,\ldots,i_k)}^{(j_1,j_2,\ldots,j_k)}(AB)\right) = \det\left(\operatorname{sub}_{w(P)}^{w(Q)}(AB)\right)$$

$$\left(\begin{array}{c}\text{since } (i_1,i_2,\ldots,i_k)=w(P) \\ \text{and } (j_1,j_2,\ldots,j_k)=w(Q)\end{array}\right),$$

we obtain

$$\det\left(\operatorname{sub}_{w(P)}^{w(Q)}(AB)\right) = \sum_{\substack{R\subseteq\{1,2,\ldots,p\};\\ |R|=k}} \det\left(\operatorname{sub}_{w(P)}^{w(R)} A\right) \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right).$$

This proves Corollary 7.251. □

Our next step towards solving Exercise 6.56 again is the following fact:

Proposition 7.252. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. For any subset I of $\{1,2,\ldots,n\}$, we let $\widetilde{I}$ denote the complement $\{1,2,\ldots,n\} \setminus I$ of I.

Let A be an $n \times n$-matrix. Let B be an $n \times m$-matrix. Let P be a subset of $\{1,2,\ldots,n\}$. Let Q be a subset of $\{1,2,\ldots,m\}$ such that $|P| = |Q|$. Then,

$$\begin{aligned}&\det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} B\right)\\ &= \sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum P + \sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \det\left(\operatorname{sub}_{w(K)}^{w(Q)}(AB)\right).\end{aligned}$$

Proof of Proposition 7.252. Set $k = |P|$. Clearly, $k = |P| = |Q|$.

If K is a subset of $\{1,2,\ldots,n\}$ satisfying $|K| = |P|$, then

$$\begin{aligned}&\det\left(\operatorname{sub}_{w(K)}^{w(Q)}(AB)\right)\\ &= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)\end{aligned} \tag{1469}$$

[662].

[662] *Proof of (1469):* Let K be a subset of $\{1,2,\ldots,n\}$ satisfying $|K| = |P|$. Then, $|K| = |P| = k$ (since $k = |P|$). Also, $|Q| = k$ (since $k = |Q|$). Hence, Corollary 7.251 (applied to n and K instead of p and P) yields

$$\det\left(\operatorname{sub}_{w(K)}^{w(Q)}(AB)\right) = \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right).$$

This proves (1469).

Now,

$$\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \underbrace{\det\left(\operatorname{sub}_{w(K)}^{w(Q)} (AB)\right)}_{\substack{=\sum\limits_{\substack{R\subseteq\{1,2,\ldots,n\};\\|R|=k}} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right)\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)\\ \text{(by (1469))}}}$$

$$= \sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \left(\sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\|R|=k}} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right)\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)\right)$$

$$= \underbrace{\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|P|}} \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\|R|=k}}}_{=\sum\limits_{\substack{R\subseteq\{1,2,\ldots,n\};\\|R|=k}} \sum\limits_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|P|}}} \underbrace{(-1)^{\sum P+\sum K}}_{\substack{=(-1)^{\sum K+\sum P}\\ \text{(since } \sum P+\sum K=\sum K+\sum P)}} \underbrace{\det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)\det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right)}_{=\det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)}\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$$

$$= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\|R|=k}} \sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|P|}} (-1)^{\sum K+\sum P} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$$

$$= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\|R|=k}} \left(\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|P|}} (-1)^{\sum K+\sum P} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)\right)\cdot\det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right). \tag{1470}$$

On the other hand, for every subset G of $\{1,2,\ldots,n\}$, we have

$$\det A = \sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\|K|=|G|}} (-1)^{\sum K+\sum G} \det\left(\operatorname{sub}_{w(K)}^{w(G)} A\right)\det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{G})} A\right). \tag{1471}$$

(Indeed, this is precisely the claim of Theorem 6.156 **(b)**, with the variables P and

Q renamed as K and G.) Applying (1471) to $G = P$, we obtain

$$\det A = \sum_{\substack{K \subseteq \{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum K + \sum P} \det\left(\operatorname{sub}_{w(K)}^{w(P)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right). \tag{1472}$$

For any two objects i and j, we define $\delta_{i,j}$ to be the element $\begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$ of $\mathbb{K}$. If R is a subset of $\{1, 2, \ldots, n\}$ satisfying $|R| = k$, then

$$\sum_{\substack{K \subseteq \{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum K + \sum P} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) = \delta_{R,P} \det A \tag{1473}$$

[663].

[663] *Proof of (1473):* Let R be a subset of $\{1, 2, \ldots, n\}$ satisfying $|R| = k$. We must prove (1473).

We have

$$\underbrace{\delta_{P,P}}_{\substack{=1\\ (\text{since } P=P)}} \det A = \det A = \sum_{\substack{K \subseteq \{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum K + \sum P} \det\left(\operatorname{sub}_{w(K)}^{w(P)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)$$

(by (1471), applied to $G = P$). In other words, (1473) holds if $R = P$. Hence, for the rest of our proof of (1473), we WLOG assume that $R \neq P$. Thus, $\delta_{R,P} = 0$.

We have $R \neq P$ and thus $P \neq R$. Also, $|P| = k = |R|$.

For every subset G of $\{1, 2, \ldots, n\}$ satisfying $|G| = |R|$ and $G \neq R$, we have

$$0 = \sum_{\substack{K \subseteq \{1,2,\ldots,n\};\\ |K|=|G|}} (-1)^{\sum K + \sum G} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{G})} A\right). \tag{1474}$$

(Indeed, this is precisely the claim of Exercise 6.45 **(b)**, with the variables Q and P renamed as G and K.) Applying (1474) to $G = P$, we obtain

$$0 = \sum_{\substack{K \subseteq \{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum K + \sum P} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)$$

(since $|P| = |R|$ and $P \neq R$). Hence,

$$\sum_{\substack{K \subseteq \{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum K + \sum P} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) = 0 = \underbrace{0}_{=\delta_{R,P}} \det A = \delta_{R,P} \det A.$$

This proves (1473).

Now, (1470) becomes

$$\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \det\left(\operatorname{sub}_{w(K)}^{w(Q)} (AB)\right)$$

$$= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \underbrace{\left(\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum K+\sum P} \det\left(\operatorname{sub}_{w(K)}^{w(R)} A\right) \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right)\right)}_{\substack{=\delta_{R,P}\det A\\ \text{(by (1473))}}}$$

$$\cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$$

$$= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \delta_{R,P} \det A \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right). \qquad (1475)$$

But P is a subset of $\{1,2,\ldots,n\}$ and satisfies $|P| = k$. In other words, P is a subset R of $\{1,2,\ldots,n\}$ satisfying $|R| = k$. Thus, the sum $\sum\limits_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \delta_{R,P} \det A \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$ has an addend for $R = P$. If we split off this addend from this sum, then we obtain

$$\sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \delta_{R,P} \det A \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$$

$$= \underbrace{\delta_{P,P}}_{\substack{=1\\ \text{(since } P=P)}} \det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} B\right) + \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k;\\ R\neq P}} \underbrace{\delta_{R,P}}_{\substack{=0\\ \text{(since } R\neq P)}} \det A \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)$$

$$= \det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} B\right) + \underbrace{\sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k;\\ R\neq P}} 0 \det A \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right)}_{=0}$$

$$= \det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} B\right).$$

Hence, (1475) becomes

$$\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \det\left(\operatorname{sub}_{w(K)}^{w(Q)} (AB)\right)$$

$$= \sum_{\substack{R\subseteq\{1,2,\ldots,n\};\\ |R|=k}} \delta_{R,P} \det A \cdot \det\left(\operatorname{sub}_{w(R)}^{w(Q)} B\right) = \det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} B\right).$$

This proves Proposition 7.252. □

Proposition 7.252 becomes particularly simple when the matrix AB is diagonal (i.e., has all entries outside of its diagonal equal to 0):

Corollary 7.253. Let $n \in \mathbb{N}$. For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let $d_1, d_2, \ldots, d_n$ be n elements of $\mathbb{K}$. Let A and B be two $n \times n$-matrices such that $AB = \left(d_i \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Let P be a subset of $\{1, 2, \ldots, n\}$. Let Q be a subset of $\{1, 2, \ldots, n\}$ such that $|P| = |Q|$. Then,

$$\det A \cdot \det \left(\operatorname{sub}_{w(P)}^{w(Q)} B \right) = (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A \right) \prod_{i \in Q} d_i.$$

Proof of Corollary 7.253. If K is a subset of $\{1, 2, \ldots, n\}$ satisfying $|K| = |P|$, then

$$\det \left(\operatorname{sub}_{w(K)}^{w(Q)} (AB) \right) = \delta_{K,Q} \prod_{i \in K} d_i \qquad (1476)$$

[664].

Proposition 7.252 (applied to $m = n$) yields

$$\begin{aligned} &\det A \cdot \det \left(\operatorname{sub}_{w(P)}^{w(Q)} B \right) \\ &= \sum_{\substack{K \subseteq \{1,2,\ldots,n\}; \\ |K| = |P|}} (-1)^{\sum P + \sum K} \det \left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A \right) \underbrace{\det \left(\operatorname{sub}_{w(K)}^{w(Q)} (AB) \right)}_{\substack{= \delta_{K,Q} \prod_{i \in K} d_i \\ \text{(by (1476))}}} \\ &= \sum_{\substack{K \subseteq \{1,2,\ldots,n\}; \\ |K| = |P|}} (-1)^{\sum P + \sum K} \det \left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A \right) \delta_{K,Q} \prod_{i \in K} d_i. \end{aligned} \qquad (1477)$$

[664] *Proof of (1476):* Let K be a subset of $\{1, 2, \ldots, n\}$ satisfying $|K| = |P|$. Then, $|K| = |P| = |Q|$. Also, AB is the $n \times n$-matrix $\left(d_i \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (since $AB = \left(d_i \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$). Hence, Lemma 6.163 (applied to AB and K instead of D and P) yields

$$\det \left(\operatorname{sub}_{w(K)}^{w(Q)} (AB) \right) = \delta_{K,Q} \prod_{i \in K} d_i.$$

This proves (1476).

But Q is a subset of $\{1,2,\ldots,n\}$ and satisfies $|Q| = |P|$ (since $|P| = |Q|$). In other words, Q is a subset K of $\{1,2,\ldots,n\}$ satisfying $|K| = |P|$. Thus, the sum

$$\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \delta_{K,Q} \prod_{i\in K} d_i$$

has an addend for $K = Q$. If we split off this addend from this sum, then we obtain

$$\begin{aligned}
&\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \delta_{K,Q} \prod_{i\in K} d_i \\
&= (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right) \underbrace{\delta_{Q,Q}}_{\substack{=1\\ (\text{since } Q=Q)}} \prod_{i\in Q} d_i \\
&\qquad + \sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|;\\ K\neq Q}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \underbrace{\delta_{K,Q}}_{\substack{=0\\ (\text{since } K\neq Q)}} \prod_{i\in K} d_i \\
&= (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right) \prod_{i\in Q} d_i \\
&\qquad + \underbrace{\sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|;\\ K\neq Q}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) 0 \prod_{i\in K} d_i}_{=0} \\
&= (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right) \prod_{i\in Q} d_i .
\end{aligned}$$

Hence, (1477) becomes

$$\begin{aligned}
&\det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} B\right) \\
&= \sum_{\substack{K\subseteq\{1,2,\ldots,n\};\\ |K|=|P|}} (-1)^{\sum P+\sum K} \det\left(\operatorname{sub}_{w(\widetilde{K})}^{w(\widetilde{P})} A\right) \delta_{K,Q} \prod_{i\in K} d_i \\
&= (-1)^{\sum P+\sum Q} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right) \prod_{i\in Q} d_i .
\end{aligned}$$

This proves Corollary 7.253. □

Specializing Corollary 7.253 a bit further, we obtain the following:

Corollary 7.254. Let $n \in \mathbb{N}$. Let $\lambda \in \mathbb{K}$.

For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let A and B be two $n \times n$-matrices such that $AB = \lambda I_n$.

Let P be a subset of $\{1, 2, \ldots, n\}$. Let Q be a subset of $\{1, 2, \ldots, n\}$ such that $|P| = |Q|$. Then,

$$\det A \cdot \det \left(\operatorname{sub}_{w(P)}^{w(Q)} B \right) = (-1)^{\sum P + \sum Q} \lambda^{|Q|} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A \right).$$

Proof of Corollary 7.254. For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ as in Corollary 7.253. We have $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n, \ 1 \leq j \leq n}$ (by the definition of I_n). Now,

$$AB = \lambda \underbrace{I_n}_{= \left(\delta_{i,j}\right)_{1 \leq i \leq n, \ 1 \leq j \leq n}} = \lambda \left(\delta_{i,j}\right)_{1 \leq i \leq n, \ 1 \leq j \leq n} = \left(\lambda \delta_{i,j}\right)_{1 \leq i \leq n, \ 1 \leq j \leq n}.$$

Hence, Corollary 7.253 (applied to $d_k = \lambda$) yields

$$\begin{aligned} \det A \cdot \det \left(\operatorname{sub}_{w(P)}^{w(Q)} B \right) &= (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A \right) \underbrace{\prod_{i \in Q} \lambda}_{= \lambda^{|Q|}} \\ &= (-1)^{\sum P + \sum Q} \lambda^{|Q|} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A \right). \end{aligned}$$

This proves Corollary 7.254. □

Finally, we obtain Exercise 6.56 easily by setting $\lambda = 1$ in Corollary 7.254:

Second solution to Exercise 6.56. The matrix $A \in \mathbb{K}^{n \times n}$ is invertible. Hence, the matrix $A^{-1} \in \mathbb{K}^{n \times n}$ is well-defined. Theorem 6.23 (applied to $B = A^{-1}$) yields $\det \left(AA^{-1} \right) = \det A \cdot \det \left(A^{-1} \right)$. Thus, $\det A \cdot \det \left(A^{-1} \right) = \det \left(\underbrace{AA^{-1}}_{= I_n} \right) = \det \left(I_n \right) = 1$.

We have $A^{-1} A = I_n = 1 \cdot I_n$. Hence, Corollary 7.254 (applied to A^{-1}, A and 1 instead of A, B and λ) yields

$$\begin{aligned} \det \left(A^{-1} \right) \cdot \det \left(\operatorname{sub}_{w(P)}^{w(Q)} A \right) &= (-1)^{\sum P + \sum Q} \underbrace{1^{|Q|}}_{= 1} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} \left(A^{-1} \right) \right) \\ &= (-1)^{\sum P + \sum Q} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} \left(A^{-1} \right) \right). \end{aligned}$$

Multiplying both sides of this equality by $\det A$, we obtain

$$\det A \cdot \det\left(A^{-1}\right) \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) = \det A \cdot (-1)^{\sum P + \sum Q} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})}\left(A^{-1}\right)\right).$$

Comparing this with

$$\underbrace{\det A \cdot \det\left(A^{-1}\right)}_{=1} \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) = \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right),$$

we obtain

$$\begin{aligned}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right) &= \det A \cdot (-1)^{\sum P + \sum Q} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})}\left(A^{-1}\right)\right) \\ &= (-1)^{\sum P + \sum Q} \det A \cdot \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})}\left(A^{-1}\right)\right).\end{aligned}$$

Thus, Exercise 6.56 is solved again. □

7.124.3. Addendum

As an easy consequence of our Second solution to Exercise 6.56, we can obtain the following fact:

Corollary 7.255. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let A be an $n \times n$-matrix.

Let P and Q be two subsets of $\{1, 2, \ldots, n\}$ such that $|P| = |Q|$. Then,

$$\det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)}(\operatorname{adj} A)\right) = (-1)^{\sum P + \sum Q} (\det A)^{|Q|} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right).$$

Proof of Corollary 7.255. Theorem 6.100 yields $A \cdot \operatorname{adj} A = \operatorname{adj} A \cdot A = \det A \cdot I_n$. Hence, Corollary 7.254 (applied to $B = \operatorname{adj} A$ and $\lambda = \det A$) yields

$$\det A \cdot \det\left(\operatorname{sub}_{w(P)}^{w(Q)}(\operatorname{adj} A)\right) = (-1)^{\sum P + \sum Q} (\det A)^{|Q|} \det\left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A\right).$$

This proves Corollary 7.255. □

Note that we could have also deduced Corollary 7.255 from the First solution to Exercise 6.56 (but this would have been more difficult, since we would have to slightly generalize our argument).

It is possible to strengthen Corollary 7.255 in the case when $|P| = |Q| \geq 1$ as follows:

Corollary 7.256. Let $n \in \mathbb{N}$. For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I.

Let A be an $n \times n$-matrix.

Let P and Q be two subsets of $\{1, 2, \ldots, n\}$ such that $|P| = |Q| \geq 1$. Then,

$$\det \left(\operatorname{sub}_{w(P)}^{w(Q)} (\operatorname{adj} A) \right) = (-1)^{\sum P + \sum Q} (\det A)^{|Q|-1} \det \left(\operatorname{sub}_{w(\widetilde{Q})}^{w(\widetilde{P})} A \right).$$

Loosely speaking, the claim of Corollary 7.256 is obtained from that of Corollary 7.255 by cancelling $\det A$. However, it is not immediately clear that this cancellation is allowed (for instance, $\det A$ could be 0, or could be a nonzero non-cancellable element). There are ways to justify this cancellation in full generality; however, these are not in the scope of these notes.

7.125. Solution to Exercise 6.57

Throughout this section, we shall use the notations introduced in Definition 6.78, in Definition 6.128 and in Definition 6.153. Also, whenever m is an integer, we shall use the notation $[m]$ for the set $\{1, 2, \ldots, m\}$. Furthermore, we shall use the following notations:

Definition 7.257. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $B \in \mathbb{K}^{n \times m}$. Let I be a subset of $\{1, 2, \ldots, m\}$.

(a) Let $B_{\bullet, \sim I}$ denote the $n \times (m - |I|)$-matrix whose columns are $B_{\bullet, j_1}, B_{\bullet, j_2}, \ldots, B_{\bullet, j_h}$ (from left to right), where $(j_1, j_2, \ldots, j_h) = w(\{1, 2, \ldots, m\} \setminus I)$. (We will see that this is well-defined in Lemma 7.258 **(a)** below.)

(b) Let $p \in \mathbb{N}$. Let $A \in \mathbb{K}^{n \times p}$. Let $(A \mid B_{\bullet, I})$ denote the $n \times (p + |I|)$-matrix whose columns are $\underbrace{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}}_{\text{the columns of } A}, B_{\bullet, i_1}, B_{\bullet, i_2}, \ldots, B_{\bullet, i_\ell}$ (from left to right), where $(i_1, i_2, \ldots, i_\ell) = w(I)$. (We will see that this is well-defined in Lemma 7.259 **(a)** below.)

This definition agrees with the notations defined in Exercise 6.57, for the following reasons:

- The notation $B_{\bullet, \sim I}$ introduced in Definition 7.257 **(a)** generalizes the notation $B_{\bullet, \sim I}$ defined in Exercise 6.57. (Indeed, the former becomes the latter when we set $m = n + k$.)
- The notation $(A \mid B_{\bullet, I})$ introduced in Definition 7.257 **(b)** generalizes the notation $(A \mid B_{\bullet, I})$ defined in Exercise 6.57. (Indeed, the former becomes the latter when we set $m = n + k$ and $p = n - k$.)

We shall now state and prove two lemmas which show that the matrices $B_{\bullet,\sim I}$ and $(A \mid B_{\bullet,I})$ from Definition 7.257 are well-defined, and express these matrices in more familiar terms (which will be helpful when we come to the solution of Exercise 6.57). Both lemmas are fairly obvious when you draw the matrices; but the rigorous proofs are rather tedious.

Lemma 7.258. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $B \in \mathbb{K}^{n\times m}$. Let I be a subset of $\{1,2,\ldots,m\}$. Then:

(a) The matrix $B_{\bullet,\sim I}$ in Definition 7.257 **(a)** is a well-defined $n \times (m-|I|)$-matrix.

(b) We have

$$B_{\bullet,\sim I} = \operatorname{sub}_{w([n])}^{w([m]\setminus I)} B.$$

(To make sense of the right-hand side of this equality, we recall that $[n] = \{1,2,\ldots,n\}$ and $[m] = \{1,2,\ldots,m\}$, and that $\operatorname{sub}_{w([n])}^{w([m]\setminus I)} B$ is defined as in Definition 6.153.)

Proof of Lemma 7.258. From $I \subseteq \{1,2,\ldots,m\}$, we obtain

$$|\{1,2,\ldots,m\} \setminus I| = \underbrace{|\{1,2,\ldots,m\}|}_{=m} - |I| = m - |I|.$$

Define a list $(j_1, j_2, \ldots, j_h)$ by $(j_1, j_2, \ldots, j_h) = w(\{1,2,\ldots,m\} \setminus I)$. Then, the definition of $B_{\bullet,\sim I}$ says that $B_{\bullet,\sim I}$ is the $n \times (m-|I|)$-matrix whose columns are $B_{\bullet,j_1}, B_{\bullet,j_2}, \ldots, B_{\bullet,j_h}$ (from left to right). In order to check that this is well-defined, we must verify that there really exists such an $n \times (m-|I|)$-matrix.

We have $(j_1, j_2, \ldots, j_h) = w(\{1,2,\ldots,m\} \setminus I)$. In other words, $(j_1, j_2, \ldots, j_h)$ is the list of all elements of $\{1,2,\ldots,m\} \setminus I$ in increasing order (with no repetitions) (because this is how $w(\{1,2,\ldots,m\} \setminus I)$ is defined). Hence, the length h of this list equals $|\{1,2,\ldots,m\} \setminus I|$. Thus, $h = |\{1,2,\ldots,m\} \setminus I| = m - |I|$.

Recall that $(j_1, j_2, \ldots, j_h) = w(\{1,2,\ldots,m\} \setminus I)$. Hence, $j_1, j_2, \ldots, j_h$ are h elements of $\{1,2,\ldots,m\} \setminus I$, and thus are elements of $\{1,2,\ldots,m\}$. Hence, $B_{\bullet,j_1}, B_{\bullet,j_2}, \ldots, B_{\bullet,j_h}$ are h column vectors with n entries each. Hence, there exists an $n \times h$-matrix whose columns are $B_{\bullet,j_1}, B_{\bullet,j_2}, \ldots, B_{\bullet,j_h}$ (from left to right). In view of $h = m - |I|$, this rewrites as follows: There exists an $n \times (m-|I|)$-matrix whose columns are $B_{\bullet,j_1}, B_{\bullet,j_2}, \ldots, B_{\bullet,j_h}$ (from left to right). In other words, the matrix $B_{\bullet,\sim I}$ in Definition 7.257 **(a)** is a well-defined $n \times (m-|I|)$-matrix. This proves Lemma 7.258 **(a)**.

(b) Write the $n \times m$-matrix B in the form $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$. The list $w([n])$ is defined as the list of all elements of $[n]$ in increasing order (with no repetitions), and thus is simply the list $(1,2,\ldots,n)$; in other words, $w([n]) = (1,2,\ldots,n)$. Moreover,

$w\left(\underbrace{[m]}_{=\{1,2,\ldots,m\}} \setminus I\right) = w\left(\{1,2,\ldots,m\} \setminus I\right) = (j_1, j_2, \ldots, j_h)$. Now,

$$\begin{aligned} \operatorname{sub}_{w([n])}^{w([m]\setminus I)} B &= \operatorname{sub}_{(1,2,\ldots,n)}^{(j_1,j_2,\ldots,j_h)} B \qquad \left(\begin{array}{c}\text{since } w([n]) = (1,2,\ldots,n) \\ \text{and } w([m]\setminus I) = (j_1,j_2,\ldots,j_h)\end{array}\right) \\ &= \operatorname{sub}_{1,2,\ldots,n}^{j_1,j_2,\ldots,j_h} B = \left(b_{x,j_y}\right)_{1\leq x\leq n,\ 1\leq y\leq h} \end{aligned} \tag{1478}$$

(by the definition of $\operatorname{sub}_{1,2,\ldots,n}^{j_1,j_2,\ldots,j_h} B$, since $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$).

On the other hand, the definition of $B_{\bullet,\sim I}$ shows that the columns of the matrix $B_{\bullet,\sim I}$ are $B_{\bullet,j_1}, B_{\bullet,j_2}, \ldots, B_{\bullet,j_h}$ (from left to right). Thus, for each $y \in \{1,2,\ldots,h\}$, we have

$$(\text{the } y\text{-th column of the matrix } B_{\bullet,\sim I}) = B_{\bullet,j_y}. \tag{1479}$$

In particular, the matrix $B_{\bullet,\sim I}$ has h columns, each of which is a column vector with n entries; thus, $B_{\bullet,\sim I}$ is an $n \times h$-matrix. For each $x \in \{1,2,\ldots,n\}$ and $y \in \{1,2,\ldots,h\}$, we have

$$\begin{aligned} &(\text{the } (x,y)\text{-th entry of the matrix } B_{\bullet,\sim I}) \\ &= (\text{the } x\text{-th entry of the } y\text{-th column of the matrix } B_{\bullet,\sim I}) \\ &= \left(\text{the } x\text{-th entry of } B_{\bullet,j_y}\right) \qquad (\text{by (1479)}) \\ &= \left(\text{the } (x,j_y)\text{-th entry of } B\right) \qquad \left(\text{since } B_{\bullet,j_y} \text{ is the } j_y\text{-th column of } B\right) \\ &= b_{x,j_y} \qquad \left(\text{since } B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}\right). \end{aligned}$$

Hence, $B_{\bullet,\sim I} = \left(b_{x,j_y}\right)_{1\leq x\leq n,\ 1\leq y\leq h}$. Comparing this with (1478), we obtain $B_{\bullet,\sim I} = \operatorname{sub}_{w([n])}^{w([m]\setminus I)} B$. This proves Lemma 7.258 **(b)**. □

Lemma 7.259. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let $B \in \mathbb{K}^{n\times m}$. Let I be a subset of $\{1,2,\ldots,m\}$. Let $A \in \mathbb{K}^{n\times p}$.

(a) The matrix $(A \mid B_{\bullet,I})$ in Definition 7.257 **(b)** is a well-defined $n \times (p + |I|)$-matrix.

(b) Let P be any subset of $\{1,2,\ldots,n\}$. Let Q be the set $\{1,2,\ldots,p\}$. Let R be the set $\{p+1, p+2, \ldots, p+|I|\}$. Then,

$$\operatorname{sub}_{w(P)}^{w(Q)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{w(P)}^{w(Q)} A$$

and

$$\operatorname{sub}_{w(P)}^{w(R)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{w(P)}^{w(I)} B.$$

Proof of Lemma 7.259. Define a list $(i_1, i_2, \ldots, i_\ell)$ by $(i_1, i_2, \ldots, i_\ell) = w(I)$. Then, the definition of $(A \mid B_{\bullet,I})$ says that $(A \mid B_{\bullet,I})$ is the $n \times (p + |I|)$-matrix whose columns are $\underbrace{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}}_{\text{the columns of } A}, B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ (from left to right). In order to check that this is well-defined, we must verify that there really exists such an $n \times (p + |I|)$-matrix.

We know that $A \in \mathbb{K}^{n \times p}$. Hence, $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}$ are p column vectors in $\mathbb{K}^{n \times 1}$: namely, the columns of A.

We have $(i_1, i_2, \ldots, i_\ell) = w(I)$. In other words, $(i_1, i_2, \ldots, i_\ell)$ is the list of all elements of I in increasing order (with no repetition) (because this is how $w(I)$ is defined). Thus, the length ℓ of this list equals $|I|$. In other words, $\ell = |I|$.

Also, $i_1, i_2, \ldots, i_\ell$ are elements of I (since $(i_1, i_2, \ldots, i_\ell)$ is the list of all elements of I in increasing order), and thus are elements of $\{1, 2, \ldots, m\}$ (since I is a subset of $\{1, 2, \ldots, m\}$). Hence, $B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ are ℓ column vectors in $\mathbb{K}^{n \times 1}$ (since $B \in \mathbb{K}^{n \times m}$).

We now conclude that $\underbrace{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}}_{\text{the columns of } A}, B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ are $p + \ell$ column vectors in $\mathbb{K}^{n \times 1}$ (because $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}$ are p column vectors in $\mathbb{K}^{n \times 1}$, and because $B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ are ℓ column vectors in $\mathbb{K}^{n \times 1}$). Thus, there exists an $n \times (p + \ell)$-matrix whose columns are $\underbrace{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}}_{\text{the columns of } A}, B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ (from left to right). In view of $\ell = |I|$, this rewrites as follows: There exists an $n \times (p + |I|)$-matrix whose columns are $\underbrace{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}}_{\text{the columns of } A}, B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ (from left to right). In other words, the matrix $(A \mid B_{\bullet,I})$ in Definition 7.257 **(b)** is a well-defined $n \times (p + |I|)$-matrix. This proves Lemma 7.259 **(a)**.

(b) The definition of R yields $R = \{p + 1, p + 2, \ldots, p + |I|\} \subseteq \{1, 2, \ldots, p + |I|\}$ (since $p \geq 0$).

The definition of Q yields $Q = \{1, 2, \ldots, p\} \subseteq \{1, 2, \ldots, p + |I|\}$ (since $|I| \geq 0$).

Recall that $(A \mid B_{\bullet,I})$ is an $n \times (p + |I|)$-matrix. In other words, $(A \mid B_{\bullet,I})$ is an $n \times (p + \ell)$-matrix (since $\ell = |I|$). Write this $n \times (p + \ell)$-matrix $(A \mid B_{\bullet,I})$ in the form $(A \mid B_{\bullet,I}) = (c_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq p+\ell}$.

Write the $n \times p$-matrix A in the form $A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq p}$.

Write the $n \times m$-matrix B in the form $B = (b_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq m}$.

The columns of the matrix $(A \mid B_{\bullet,I})$ (from left to right) are $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,p}, B_{\bullet,i_1}, B_{\bullet,i_2}, \ldots, B_{\bullet,i_\ell}$ (by the definition of $(A \mid B_{\bullet,I})$). Thus, for each $g \in \{1, 2, \ldots, p + \ell\}$, we have

$$(\text{the } g\text{-th column of the matrix } (A \mid B_{\bullet,I})) = \begin{cases} A_{\bullet,g}, & \text{if } g \leq p; \\ B_{\bullet,i_{g-p}}, & \text{if } g > p \end{cases}. \tag{1480}$$

For each $x \in \{1, 2, \ldots, n\}$ and $g \in \{1, 2, \ldots, p\}$, we have

$$c_{x,g} = a_{x,g} \tag{1481}$$

[665].

For each $x \in \{1, 2, \ldots, n\}$ and $g \in \{p+1, p+2, \ldots, p+\ell\}$, we have

$$c_{x,g} = b_{x,i_{g-p}} \tag{1483}$$

[666]. Hence, for each $x \in \{1, 2, \ldots, n\}$ and $y \in \{1, 2, \ldots, \ell\}$, we have

$$c_{x,p+y} = b_{x,i_y} \tag{1485}$$

[665] *Proof of (1481):* Let $x \in \{1, 2, \ldots, n\}$ and $g \in \{1, 2, \ldots, p\}$. Thus, $g \leq p$. Also, $g \in \{1, 2, \ldots, p\} \subseteq \{1, 2, \ldots, p+\ell\}$ (since $\ell \geq 0$). Hence, (1480) yields

$$\begin{aligned} &(\text{the } g\text{-th column of the matrix } (A \mid B_{\bullet,I})) \\ &= \begin{cases} A_{\bullet,g}, & \text{if } g \leq p; \\ B_{\bullet,i_{g-p}}, & \text{if } g > p \end{cases} = A_{\bullet,g} \qquad (\text{since } g \leq p) \\ &= (\text{the } g\text{-th column of the matrix } A) \end{aligned} \tag{1482}$$

(since $A_{\bullet,g}$ is defined to be the g-th column of the matrix A). But

$$(\text{the } (x,g)\text{-th entry of the matrix } (A \mid B_{\bullet,I})) = c_{x,g}$$

(since $(A \mid B_{\bullet,I}) = (c_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq p+\ell}$). Comparing this with

$$\begin{aligned} &(\text{the } (x,g)\text{-th entry of the matrix } (A \mid B_{\bullet,I})) \\ &= \left(\text{the } x\text{-th entry of } \underbrace{\text{the } g\text{-th column of the matrix } (A \mid B_{\bullet,I})}_{\substack{=(\text{the } g\text{-th column of the matrix } A) \\ (\text{by (1482)})}} \right) \\ &= (\text{the } x\text{-th entry of the } g\text{-th column of the matrix } A) \\ &= (\text{the } (x,g)\text{-th entry of the matrix } A) = a_{x,g} \qquad \left(\text{since } A = (a_{i,j})_{1 \leq i \leq n,\ 1 \leq j \leq p}\right), \end{aligned}$$

we obtain $c_{x,g} = a_{x,g}$. This proves (1481).

[666] *Proof of (1483):* Let $x \in \{1, 2, \ldots, n\}$ and $g \in \{p+1, p+2, \ldots, p+\ell\}$. Thus, $g \geq p+1 > p$. Also, $g \in \{p+1, p+2, \ldots, p+\ell\} \subseteq \{1, 2, \ldots, p+\ell\}$. Hence, (1480) yields

$$\begin{aligned} &(\text{the } g\text{-th column of the matrix } (A \mid B_{\bullet,I})) \\ &= \begin{cases} A_{\bullet,g}, & \text{if } g \leq p; \\ B_{\bullet,i_{g-p}}, & \text{if } g > p \end{cases} = B_{\bullet,i_{g-p}} \qquad (\text{since } g > p) \\ &= \left(\text{the } i_{g-p}\text{-th column of the matrix } B\right) \end{aligned} \tag{1484}$$

(since $B_{\bullet,i_{g-p}}$ is defined to be the i_{g-p}-th column of the matrix B). But

$$(\text{the } (x,g)\text{-th entry of the matrix } (A \mid B_{\bullet,I})) = c_{x,g}$$

[667].

Now, $P \subseteq \{1,2,\ldots,n\}$, whereas $Q \subseteq \{1,2,\ldots,p+|I|\}$. Thus, the matrix $\operatorname{sub}_{w(P)}^{w(Q)}(A \mid B_{\bullet,I})$ is well-defined (since $(A \mid B_{\bullet,I})$ is an $n \times (p+|I|)$-matrix). Also, $P \subseteq \{1,2,\ldots,n\}$, whereas $Q \subseteq \{1,2,\ldots,p\}$ (since $Q = \{1,2,\ldots,p\}$). Hence, the matrix $\operatorname{sub}_{w(P)}^{w(Q)} A$ is well-defined (since A is an $n \times p$-matrix).

Furthermore, $P \subseteq \{1,2,\ldots,n\}$, whereas $R \subseteq \{1,2,\ldots,p+|I|\}$. Thus, the matrix $\operatorname{sub}_{w(P)}^{w(R)}(A \mid B_{\bullet,I})$ is well-defined (since $(A \mid B_{\bullet,I})$ is an $n \times (p+|I|)$-matrix). Also, $P \subseteq \{1,2,\ldots,n\}$, whereas $I \subseteq \{1,2,\ldots,m\}$. Hence, the matrix $\operatorname{sub}_{w(P)}^{w(I)} B$ is well-defined (since B is an $n \times m$-matrix).

We have $Q = \{1,2,\ldots,p\}$, and thus $w(Q) = (1,2,\ldots,p)$ (by the definition of $w(Q)$).

Also, $R = \{p+1,p+2,\ldots,p+|I|\} = \{p+1,p+2,\ldots,p+\ell\}$ (since $|I| = \ell$), and thus $w(R) = (p+1,p+2,\ldots,p+\ell)$ (by the definition of $w(R)$).

Write the list $w(P)$ in the form $w(P) = (p_1,p_2,\ldots,p_h)$. Thus, $p_1,p_2,\ldots,p_h$ are all the elements of P (since $w(P)$ is the list of all elements of P in increasing order). Hence, $p_1,p_2,\ldots,p_h$ are elements of P, and therefore are elements of $\{1,2,\ldots,n\}$ (since $P \subseteq \{1,2,\ldots,n\}$). In other words, $p_x \in \{1,2,\ldots,n\}$ for each $x \in \{1,2,\ldots,h\}$. Therefore, for each $x \in \{1,2,\ldots,h\}$ and $y \in \{1,2,\ldots,p\}$, we have $c_{p_x,y} = a_{p_x,y}$ (by (1481) (applied to p_x and y instead of x and g)). In other words, we have

$$\left(c_{p_x,y}\right)_{1\leq x\leq h,\ 1\leq y\leq p} = \left(a_{p_x,y}\right)_{1\leq x\leq h,\ 1\leq y\leq p}. \tag{1486}$$

Recall that $p_x \in \{1,2,\ldots,n\}$ for each $x \in \{1,2,\ldots,h\}$. Therefore, for each $x \in \{1,2,\ldots,h\}$ and $y \in \{1,2,\ldots,\ell\}$, we have $c_{p_x,p+y} = b_{p_x,i_y}$ (by (1485) (applied to p_x instead of x)). In other words, we have

$$\left(c_{p_x,p+y}\right)_{1\leq x\leq h,\ 1\leq y\leq \ell} = \left(b_{p_x,i_y}\right)_{1\leq x\leq h,\ 1\leq y\leq \ell}. \tag{1487}$$

(since $(A \mid B_{\bullet,I}) = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p+\ell}$). Comparing this with

$$\begin{aligned}
&\left(\text{the } (x,g)\text{-th entry of the matrix } (A \mid B_{\bullet,I})\right)\\
&= \left(\text{the } x\text{-th entry of } \underbrace{\text{the } g\text{-th column of the matrix } (A \mid B_{\bullet,I})}_{\substack{=\left(\text{the } i_{g-p}\text{-th column of the matrix } B\right)\\ \text{(by (1484))}}}\right)\\
&= \left(\text{the } x\text{-th entry of the } i_{g-p}\text{-th column of the matrix } B\right)\\
&= \left(\text{the } (x,i_{g-p})\text{-th entry of the matrix } B\right) = b_{x,i_{g-p}} \qquad \left(\text{since } B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}\right),
\end{aligned}$$

we obtain $c_{x,g} = b_{x,i_{g-p}}$. This proves (1483).

[667] *Proof of (1485):* Let $x \in \{1,2,\ldots,n\}$ and $y \in \{1,2,\ldots,\ell\}$. From $y \in \{1,2,\ldots,\ell\}$, we obtain $p+y \in \{p+1,p+2,\ldots,p+\ell\}$. Hence, (1483) (applied to $g = p+y$) yields $c_{x,p+y} = b_{x,i_{p+y-p}} = b_{x,i_y}$ (since $p+y-p = y$). This proves (1485).

Using $w(P) = (p_1, p_2, \ldots, p_h)$ and $w(Q) = (1, 2, \ldots, p)$, we see that

$$\begin{aligned}
\operatorname{sub}_{w(P)}^{w(Q)} (A \mid B_{\bullet,I}) &= \operatorname{sub}_{(p_1,p_2,\ldots,p_h)}^{(1,2,\ldots,p)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{1,2,\ldots,p} (A \mid B_{\bullet,I}) \\
&= \left(c_{p_x,y}\right)_{1\leq x\leq h,\ 1\leq y\leq p} \\
&\qquad \left(\begin{array}{c}\text{by the definition of } \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{1,2,\ldots,p} (A \mid B_{\bullet,I}), \\ \text{since } (A \mid B_{\bullet,I}) = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p+\ell}\end{array}\right) \\
&= \left(a_{p_x,y}\right)_{1\leq x\leq h,\ 1\leq y\leq p} \qquad \text{(by (1486))}.
\end{aligned}$$

Comparing this with

$$\begin{aligned}
\operatorname{sub}_{w(P)}^{w(Q)} A &= \operatorname{sub}_{(p_1,p_2,\ldots,p_h)}^{(1,2,\ldots,p)} A \qquad \left(\begin{array}{c}\text{since } w(P) = (p_1, p_2, \ldots, p_h) \\ \text{and } w(Q) = (1, 2, \ldots, p)\end{array}\right) \\
&= \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{1,2,\ldots,p} A = \left(a_{p_x,y}\right)_{1\leq x\leq h,\ 1\leq y\leq p} \\
&\qquad \left(\begin{array}{c}\text{by the definition of } \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{1,2,\ldots,p} A, \\ \text{since } A = \left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p}\end{array}\right),
\end{aligned}$$

we obtain

$$\operatorname{sub}_{w(P)}^{w(Q)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{w(P)}^{w(Q)} A.$$

Using $w(P) = (p_1, p_2, \ldots, p_h)$ and $w(R) = (p+1, p+2, \ldots, p+\ell)$, we find

$$\begin{aligned}
\operatorname{sub}_{w(P)}^{w(R)} (A \mid B_{\bullet,I}) &= \operatorname{sub}_{(p_1,p_2,\ldots,p_h)}^{(p+1,p+2,\ldots,p+\ell)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{p+1,p+2,\ldots,p+\ell} (A \mid B_{\bullet,I}) \\
&= \left(c_{p_x,p+y}\right)_{1\leq x\leq h,\ 1\leq y\leq \ell} \\
&\qquad \left(\begin{array}{c}\text{by the definition of } \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{p+1,p+2,\ldots,p+\ell} (A \mid B_{\bullet,I}), \\ \text{since } (A \mid B_{\bullet,I}) = \left(c_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq p+\ell}\end{array}\right) \\
&= \left(b_{p_x,i_y}\right)_{1\leq x\leq h,\ 1\leq y\leq \ell} \qquad \text{(by (1487))}.
\end{aligned}$$

Comparing this with

$$\begin{aligned}
\operatorname{sub}_{w(P)}^{w(I)} B &= \operatorname{sub}_{(p_1,p_2,\ldots,p_h)}^{(i_1,i_2,\ldots,i_\ell)} B \qquad \left(\begin{array}{c}\text{since } w(I) = (i_1, i_2, \ldots, i_\ell) \\ \text{and } w(P) = (p_1, p_2, \ldots, p_h)\end{array}\right) \\
&= \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{i_1,i_2,\ldots,i_\ell} B = \left(b_{p_x,i_y}\right)_{1\leq x\leq h,\ 1\leq y\leq \ell} \\
&\qquad \left(\begin{array}{c}\text{by the definition of } \operatorname{sub}_{p_1,p_2,\ldots,p_h}^{i_1,i_2,\ldots,i_\ell} B, \\ \text{since } B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}\end{array}\right),
\end{aligned}$$

we obtain

$$\operatorname{sub}_{w(P)}^{w(R)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{w(P)}^{w(I)} B.$$

Thus, the proof of Lemma 7.259 **(b)** is complete. □

Our next lemma is just a restatement of Exercise 6.46 **(a)** using more opportune notations:

Lemma 7.260. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let J and K be two subsets of $\{1, 2, \ldots, m\}$ satisfying $|J| + |K| = n$ and $J \cap K \neq \varnothing$. Let $A \in \mathbb{K}^{m \times n}$. Then,

$$\sum_{\substack{I \subseteq \{1,2,\ldots,n\}; \\ |I| = |J|}} (-1)^{\sum I} \det \left(\operatorname{sub}_{w(J)}^{w(I)} A \right) \det \left(\operatorname{sub}_{w(K)}^{w([n] \setminus I)} A \right) = 0.$$

Proof of Lemma 7.260. We have $[n] = \{1, 2, \ldots, n\}$ (by the definition of $[n]$).

For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. Thus, for any subset I of $\{1, 2, \ldots, n\}$, we have

$$\widetilde{I} = \underbrace{\{1, 2, \ldots, n\}}_{= [n]} \setminus I = [n] \setminus I. \tag{1488}$$

Now, Exercise 6.46 **(a)** yields

$$\begin{aligned}
0 &= \sum_{\substack{Q \subseteq \{1,2,\ldots,n\}; \\ |Q| = |J|}} (-1)^{\sum Q} \det \left(\operatorname{sub}_{w(J)}^{w(Q)} A \right) \det \left(\operatorname{sub}_{w(K)}^{w(\widetilde{Q})} A \right) \\
&= \sum_{\substack{I \subseteq \{1,2,\ldots,n\}; \\ |I| = |J|}} (-1)^{\sum I} \det \left(\operatorname{sub}_{w(J)}^{w(I)} A \right) \underbrace{\det \left(\operatorname{sub}_{w(K)}^{w(\widetilde{I})} A \right)}_{\substack{= \det \left(\operatorname{sub}_{w(K)}^{w([n] \setminus I)} A \right) \\ \text{(since } \widetilde{I} = [n] \setminus I \\ \text{(by (1488)))}}} \\
&\qquad \left(\text{here, we have renamed the summation index } Q \text{ as } I \right) \\
&= \sum_{\substack{I \subseteq \{1,2,\ldots,n\}; \\ |I| = |J|}} (-1)^{\sum I} \det \left(\operatorname{sub}_{w(J)}^{w(I)} A \right) \det \left(\operatorname{sub}_{w(K)}^{w([n] \setminus I)} A \right).
\end{aligned}$$

This proves Lemma 7.260. □

We can now solve Exercise 6.57:

Solution to Exercise 6.57. Combining $k \in \mathbb{N}$ (since k is a positive integer) with $k \leq n$, we obtain $k \in \{0, 1, \ldots, n\}$. Hence, $n - k \in \{0, 1, \ldots, n\}$.

Let $p = n - k$. Thus, $p = n - k \in \{0, 1, \ldots, n\} \subseteq \mathbb{N}$, so that $p \geq 0$ and therefore $p + 1 \geq 0 + 1 = 1$. Also, $p \leq n$ (since $p \in \{0, 1, \ldots, n\}$).

Let Q be the set $\{1, 2, \ldots, p\}$. Thus, $Q \subseteq \{1, 2, \ldots, n\}$ (since $p \leq n$) and $|Q| = p$.

Notice also that $Q = \{1, 2, \ldots, p\} = \{1, 2, \ldots, n - k\}$ (since $p = n - k$), so that Q is a subset of $\{1, 2, \ldots, n - k\}$. Hence, a matrix $\operatorname{sub}_{w(P)}^{w(Q)} A$ is well-defined whenever P is a subset of $\{1, 2, \ldots, n\}$ (since A is an $n \times (n - k)$-matrix).

Let R be the set $\{p+1, p+2, \ldots, n\}$. Thus, $R \subseteq \{1, 2, \ldots, n\}$ (since $p \geq 0$).

For any subset I of $\{1, 2, \ldots, n\}$, we let $\widetilde{I}$ denote the complement $\{1, 2, \ldots, n\} \setminus I$ of I. Thus,

$$\widetilde{Q} = \{1, 2, \ldots, n\} \setminus \underbrace{Q}_{=\{1,2,\ldots,p\}} = \{1, 2, \ldots, n\} \setminus \{1, 2, \ldots, p\}$$
$$= \{p+1, p+2, \ldots, n\} = R \qquad (\text{since } R = \{p+1, p+2, \ldots, n\}).$$

For any subset P of $\{1, 2, \ldots, n\}$ satisfying $|P| = |Q|$, we define an element $\gamma_P \in \mathbb{K}$ by

$$\gamma_P = (-1)^{\sum P + \sum Q} \det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right). \tag{1489}$$

Next, let us prove that if I is any subset of $\{1, 2, \ldots, n+k\}$ satisfying $|I| = k$, then the determinants $\det(A \mid B_{\bullet,I})$ and $\det(B_{\bullet,\sim I})$ are well-defined and the equality

$$\det(A \mid B_{\bullet,I}) = \sum_{\substack{P \subseteq \{1,2,\ldots,n\};\\ |P|=|Q|}} \gamma_P \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(I)} B\right) \tag{1490}$$

holds.

[*Proof:* Let I be a subset of $\{1, 2, \ldots, n+k\}$ satisfying $|I| = k$. Then, $(A \mid B_{\bullet,I})$ is an $n \times (n - k + |I|)$-matrix (by the definition of $(A \mid B_{\bullet,I})$). In other words, $(A \mid B_{\bullet,I})$ is an $n \times n$-matrix (since $n - k + \underbrace{|I|}_{=k} = n - k + k = n$). Thus, its determinant $\det(A \mid B_{\bullet,I})$ is well-defined.

Also, $B_{\bullet,\sim I}$ is an $n \times (n + k - |I|)$-matrix (by the definition of $B_{\bullet,\sim I}$). In other words, $B_{\bullet,\sim I}$ is an $n \times n$-matrix (since $n + k - \underbrace{|I|}_{=k} = n + k - k = n$). Hence, its determinant $\det(B_{\bullet,\sim I})$ is well-defined.

It remains to show that the equality (1490) holds.

We have $A \in \mathbb{K}^{n \times (n-k)} = \mathbb{K}^{n \times p}$ (since $n - k = p$). Furthermore, R is the set $\{p+1, p+2, \ldots, p+|I|\}$ [668]. Hence, Lemma 7.259 **(b)** (applied to $m = n + k$) shows that if P is any subset of $\{1, 2, \ldots, n\}$, then

$$\operatorname{sub}_{w(P)}^{w(Q)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{w(P)}^{w(Q)} A \tag{1491}$$

and

$$\operatorname{sub}_{w(P)}^{w(R)} (A \mid B_{\bullet,I}) = \operatorname{sub}_{w(P)}^{w(I)} B. \tag{1492}$$

Also, if P is any subset of $\{1, 2, \ldots, n\}$, then $\widetilde{P}$ is a subset of $\{1, 2, \ldots, n\}$ as well (since the definition of $\widetilde{P}$ yields $\widetilde{P} = \{1, 2, \ldots, n\} \setminus P \subseteq \{1, 2, \ldots, n\}$).

[668] *Proof.* Adding the equalities $p = n - k$ and $|I| = k$ together, we obtain $p + |I| = (n-k) + k = n$. Hence, $\{p+1, p+2, \ldots, p+|I|\} = \{p+1, p+2, \ldots, n\}$. Comparing this with $R = \{p+1, p+2, \ldots, n\}$, we obtain $R = \{p+1, p+2, \ldots, p+|I|\}$. In other words, R is the set $\{p+1, p+2, \ldots, p+|I|\}$.

But $(A \mid B_{\bullet,I})$ is an $n \times n$-matrix. In other words, $(A \mid B_{\bullet,I}) \in \mathbb{K}^{n\times n}$. Hence, Theorem 6.156 **(b)** (applied to $(A \mid B_{\bullet,I})$ instead of A) yields

$$\begin{aligned}
&\det(A \mid B_{\bullet,I}) \\
&= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} (-1)^{\sum P+\sum Q} \det\left(\underbrace{\operatorname{sub}_{w(P)}^{w(Q)}(A \mid B_{\bullet,I})}_{\substack{=\operatorname{sub}_{w(P)}^{w(Q)} A \\ \text{(by (1491))}}}\right) \underbrace{\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(\widetilde{Q})}(A \mid B_{\bullet,I})\right)}_{\substack{=\det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(R)}(A|B_{\bullet,I})\right) \\ \text{(since } \widetilde{Q}=R\text{)}}} \\
&= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} \underbrace{(-1)^{\sum P+\sum Q}\det\left(\operatorname{sub}_{w(P)}^{w(Q)} A\right)}_{\substack{=\gamma_P \\ \text{(by (1489))}}} \det\left(\underbrace{\operatorname{sub}_{w(\widetilde{P})}^{w(R)}(A \mid B_{\bullet,I})}_{\substack{=\operatorname{sub}_{w(\widetilde{P})}^{w(I)} B \\ \text{(by (1492)} \\ \text{(applied to } \widetilde{P} \text{ instead of } P\text{))}}}\right) \\
&= \sum_{\substack{P\subseteq\{1,2,\ldots,n\};\\ |P|=|Q|}} \gamma_P \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(I)} B\right).
\end{aligned}$$

In other words, the equality (1490) holds.

Now, forget that we fixed I. We thus have shown that if I is any subset of $\{1,2,\ldots,n+k\}$ satisfying $|I| = k$, then the determinants $\det(A \mid B_{\bullet,I})$ and $\det(B_{\bullet,\sim I})$ are well-defined and the equality (1490) holds. Qed.]

On the other hand, if P is any subset of $\{1,2,\ldots,n\}$ satisfying $|P| = |Q|$, then

$$\sum_{\substack{I\subseteq\{1,2,\ldots,n+k\};\\ |I|=k}} (-1)^{\sum I} \det\left(\operatorname{sub}_{w(\widetilde{P})}^{w(I)} B\right) \det(B_{\bullet,\sim I}) = 0. \tag{1493}$$

[*Proof of (1493):* Let P be any subset of $\{1,2,\ldots,n\}$ satisfying $|P| = |Q|$. Thus, $|P| = |Q| = p = n-k$.

Since P is a subset of $\{1,2,\ldots,n\}$, its complement $\widetilde{P}$ is a subset of $\{1,2,\ldots,n\}$ as well and satisfies

$$\left|\widetilde{P}\right| = \underbrace{|\{1,2,\ldots,n\}|}_{=n} - \underbrace{|P|}_{=n-k} = n-(n-k) = k.$$

Furthermore, $[n]$ is a subset of $\{1,2,\ldots,n\}$ (since $[n] = \{1,2,\ldots,n\}$) and satisfies $\underbrace{\left|\widetilde{P}\right|}_{=k} + \underbrace{|[n]|}_{=n} = k+n = n+k$. Also, $\left|\widetilde{P}\right| = k > 0$ (since k is a positive integer), so that

$\widetilde{P} \neq \varnothing$. But $\widetilde{P} \subseteq \{1, 2, \ldots, n\} = [n]$ and thus $\widetilde{P} \cap [n] = \widetilde{P} \neq \varnothing$.

Hence, Lemma 7.260 (applied to n, $n+k$, B, $\widetilde{P}$ and $[n]$ instead of m, n, A, J and K) yields

$$\sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = \left|\widetilde{P}\right|}} (-1)^{\sum I} \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right) \det\left(\operatorname{sub}_{w([n])}^{w([n+k] \setminus I)} B\right) = 0.$$

In view of $\left|\widetilde{P}\right| = k$, this rewrites as

$$\sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = k}} (-1)^{\sum I} \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right) \det\left(\operatorname{sub}_{w([n])}^{w([n+k] \setminus I)} B\right) = 0. \qquad (1494)$$

Now,

$$\begin{aligned} &\sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = k}} (-1)^{\sum I} \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right) \det\left(\underbrace{B_{\bullet,\sim I}}_{\substack{= \operatorname{sub}_{w([n])}^{w([n+k] \setminus I)} B \\ \text{(by Lemma 7.258 \textbf{(b)} (applied to } m = n+k))}}\right) \\ &= \sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = k}} (-1)^{\sum I} \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right) \det\left(\operatorname{sub}_{w([n])}^{w([n+k] \setminus I)} B\right) = 0 \end{aligned}$$

(by (1494)). This proves (1493).]

Now,

$$\begin{aligned} &\sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = k}} (-1)^{\sum I} \underbrace{\det\left(A \mid B_{\bullet,I}\right)}_{\substack{= \sum\limits_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} \gamma_P \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right) \\ \text{(by (1490))}}} \det\left(B_{\bullet,\sim I}\right) \\ &= \sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = k}} (-1)^{\sum I} \left(\sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} \gamma_P \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right)\right) \det\left(B_{\bullet,\sim I}\right) \\ &= \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} \gamma_P \underbrace{\sum_{\substack{I \subseteq \{1,2,\ldots,n+k\}; \\ |I| = k}} (-1)^{\sum I} \det\left(\operatorname{sub}_{w\left(\widetilde{P}\right)}^{w(I)} B\right) \det\left(B_{\bullet,\sim I}\right)}_{\substack{=0 \\ \text{(by (1493))}}} \\ &= \sum_{\substack{P \subseteq \{1,2,\ldots,n\}; \\ |P| = |Q|}} \gamma_P 0 = 0. \end{aligned} \qquad (1495)$$

Hence,

$$\sum_{\substack{I\subseteq\{1,2,\ldots,n+k\};\\ |I|=k}} \underbrace{(-1)^{\sum I+(1+2+\cdots+k)}}_{=(-1)^{\sum I}(-1)^{1+2+\cdots+k}} \det\left(A \mid B_{\bullet,I}\right)\det\left(B_{\bullet,\sim I}\right)$$
$$= (-1)^{1+2+\cdots+k} \underbrace{\sum_{\substack{I\subseteq\{1,2,\ldots,n+k\};\\ |I|=k}} (-1)^{\sum I} \det\left(A \mid B_{\bullet,I}\right)\det\left(B_{\bullet,\sim I}\right)}_{\substack{=0\\ \text{(by (1495))}}}$$
$$= 0.$$

This solves Exercise 6.57. □

7.126. Solution to Exercise 6.59

In this section, we shall use the following notation:

Definition 7.261. If B is any 1×1-matrix, then $\operatorname{ent} B$ will denote the $(1,1)$-th entry of B. (This entry is, of course, the only entry of B. Thus, the 1×1-matrix B satisfies $B = \left(\begin{array}{c}\operatorname{ent} B\end{array}\right)$.)

We can now restate Exercise 6.59 in a form that uses no abuse of notation (such as identifying 1×1-matrices with elements of $\mathbb{K}$):

Theorem 7.262. Let $n\in\mathbb{N}$. Let u be a column vector with n entries, and let v be a row vector with n entries. (Thus, uv is an $n\times n$-matrix, whereas vu is a 1×1-matrix.) Let A be an $n\times n$-matrix. Then,

$$\det\left(A+uv\right) = \det A + \operatorname{ent}\left(v\left(\operatorname{adj} A\right)u\right).$$

Before we prove this theorem, let us make some preparations. First comes a simple fact:

Proposition 7.263. Let $n\in\mathbb{N}$ and $m\in\mathbb{N}$. Let $u = \left(u_1,u_2,\ldots,u_m\right)^T \in \mathbb{K}^{m\times 1}$ and $v = \left(v_1,v_2,\ldots,v_n\right)\in\mathbb{K}^{1\times n}$. Let $B = \left(b_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} \in \mathbb{K}^{n\times m}$. Then,

$$\operatorname{ent}\left(vBu\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} u_j v_i b_{i,j}.$$

Proof of Proposition 7.263. We have

$$u = \left(u_1,u_2,\ldots,u_m\right)^T = \left(\begin{array}{c} u_1\\ u_2\\ \vdots\\ u_m\end{array}\right) = \left(u_i\right)_{1\leq i\leq m,\ 1\leq j\leq 1}$$

and

$$v = (v_1, v_2, \ldots, v_n) = (v_j)_{1\leq i\leq 1,\ 1\leq j\leq n}.$$

The definition of the product of two matrices yields

$$vB = \left(\sum_{k=1}^{n} v_k b_{k,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq m}$$

(since $v = (v_j)_{1\leq i\leq 1,\ 1\leq j\leq n}$ and $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$). Thus,

$$vB = \left(\underbrace{\sum_{k=1}^{n} v_k b_{k,j}}_{\substack{=\sum_{p=1}^{n} v_p b_{p,j} \\ \text{(here, we have renamed the} \\ \text{summation index } k \text{ as } p)}}\right)_{1\leq i\leq 1,\ 1\leq j\leq m} = \left(\sum_{p=1}^{n} v_p b_{p,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq m}.$$

Now, the definition of the product of two matrices yields

$$(vB)\,u = \left(\sum_{k=1}^{m}\left(\sum_{p=1}^{n} v_p b_{p,k}\right) u_k\right)_{1\leq i\leq 1,\ 1\leq j\leq 1} \tag{1496}$$

(since $vB = \left(\sum_{p=1}^{n} v_p b_{p,j}\right)_{1\leq i\leq 1,\ 1\leq j\leq m}$ and $u = (u_i)_{1\leq i\leq m,\ 1\leq j\leq 1}$). In particular, $(vB)\,u$ is a 1×1-matrix. The definition of $\operatorname{ent}((vB)\,u)$ shows that $\operatorname{ent}((vB)\,u)$ is the $(1,1)$-th entry of this matrix $(vB)\,u$. Thus,

$$\begin{aligned}
\operatorname{ent}((vB)\,u) &= \left(\text{the } (1,1)\text{-th entry of the matrix } \underbrace{(vB)\,u}_{\substack{=\left(\sum\limits_{k=1}^{m}\left(\sum\limits_{p=1}^{n} v_p b_{p,k}\right)u_k\right)_{1\leq i\leq 1,\ 1\leq j\leq 1} \\ \text{(by (1496))}}}\right) \\
&= \left(\text{the } (1,1)\text{-th entry of the matrix } \left(\sum_{k=1}^{m}\left(\sum_{p=1}^{n} v_p b_{p,k}\right)u_k\right)_{1\leq i\leq 1,\ 1\leq j\leq 1}\right) \\
&= \sum_{k=1}^{m}\left(\sum_{p=1}^{n} v_p b_{p,k}\right)u_k = \underbrace{\sum_{k=1}^{m}\sum_{p=1}^{n}}_{=\sum\limits_{p=1}^{n}\sum\limits_{k=1}^{m}} \underbrace{v_p b_{p,k} u_k}_{=u_k v_p b_{p,k}} \\
&= \sum_{p=1}^{n}\sum_{k=1}^{m} u_k v_p b_{p,k} = \sum_{i=1}^{n}\sum_{k=1}^{m} u_k v_i b_{i,k} \\
&\qquad \text{(here, we have renamed the summation index } p \text{ as } i\text{)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} u_j v_i b_{i,j}
\end{aligned}$$

(here, we have renamed the summation index k as j). Since $(vB)\,u = vBu$, this rewrites as

$$\operatorname{ent}(vBu) = \sum_{i=1}^{n}\sum_{j=1}^{m} u_j v_i b_{i,j}.$$

This proves Proposition 7.263. ☐

From Proposition 7.263, we can easily obtain the following:

Corollary 7.264. Let $n \in \mathbb{N}$. Let $u = (u_1, u_2, \ldots, u_n)^T \in \mathbb{K}^{n\times 1}$ and $v = (v_1, v_2, \ldots, v_n) \in \mathbb{K}^{1\times n}$. Let A be an $n \times n$-matrix. Then,

$$\operatorname{ent}(v\,(\operatorname{adj} A)\,u) = \sum_{i=1}^{n}\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right).$$

(Here, we are using the notations introduced in Definition 6.81.)

Proof of Corollary 7.264. The definition of adj A yields

$$\operatorname{adj} A = \left((-1)^{i+j}\det\left(A_{\sim j,\sim i}\right)\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Hence, Proposition 7.263 (applied to $m = n$, $B = \operatorname{adj} A$ and $b_{i,j} = (-1)^{i+j} \det\left(A_{\sim j,\sim i}\right)$) yields

$$\begin{aligned}\operatorname{ent}\left(v\left(\operatorname{adj} A\right)u\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{u_j v_i (-1)^{i+j}}_{=(-1)^{i+j}u_j v_i} \det\left(A_{\sim j,\sim i}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right).\end{aligned}$$

This proves Corollary 7.264. □

Next, we state something slightly more interesting:

Proposition 7.265. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times m}$ and $B \in \mathbb{K}^{n\times m}$. Let $u \in \mathbb{K}^{n\times 1}$ and $v \in \mathbb{K}^{n\times 1}$. Then,

$$(A \mid u)(B \mid v)^T = AB^T + uv^T.$$

(Here, we are using the notations introduced in Definition 6.132.)

Example 7.266. If we set $n = 2$, $m = 3$, $A = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \end{pmatrix}$, $B = \begin{pmatrix} c & c' & c'' \\ d & d' & d'' \end{pmatrix}$, $u = \begin{pmatrix} x \\ y \end{pmatrix}$ and $v = \begin{pmatrix} z \\ w \end{pmatrix}$, then Proposition 7.265 says that

$$\begin{aligned}&\begin{pmatrix} a & a' & a'' & x \\ b & b' & b'' & y \end{pmatrix}\begin{pmatrix} c & c' & c'' & z \\ d & d' & d'' & w \end{pmatrix}^T \\ &= \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \end{pmatrix}\begin{pmatrix} c & c' & c'' \\ d & d' & d'' \end{pmatrix}^T + \begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} z \\ w \end{pmatrix}^T.\end{aligned}$$

First proof of Proposition 7.265 (sketched). We can apply Exercise 6.28 to n, 0, m, 1, n, 0, A, u, $0_{0\times m}$, $0_{0\times 1}$, B^T, $0_{m\times 0}$, v^T and $0_{1\times 0}$ instead of n, n', m, m', ℓ, ℓ', A, B, C, D, A', B', C' and D'. As a result, we obtain

$$\begin{pmatrix} A & u \\ 0_{0\times m} & 0_{0\times 1} \end{pmatrix}\begin{pmatrix} B^T & 0_{m\times 0} \\ v^T & 0_{1\times 0} \end{pmatrix} = \begin{pmatrix} AB^T + uv^T & A0_{m\times 0} + u0_{1\times 0} \\ 0_{0\times m}B^T + 0_{0\times 1}v^T & 0_{0\times m}0_{m\times 0} + 0_{0\times 1}0_{1\times 0} \end{pmatrix}$$

(where we are using the notations introduced in Definition 6.89). In view of the equalities

$$\begin{pmatrix} A & u \\ 0_{0\times m} & 0_{0\times 1} \end{pmatrix} = (A \mid u) \qquad \text{(this is rather obvious)},$$

$$\begin{pmatrix} B^T & 0_{m\times 0} \\ v^T & 0_{1\times 0} \end{pmatrix} = (B \mid v)^T \qquad \text{(this is easy to see)}$$

and

$$\begin{pmatrix} AB^T + uv^T & A0_{m\times 0} + u0_{1\times 0} \\ 0_{0\times m}B^T + 0_{0\times 1}v^T & 0_{0\times m}0_{m\times 0} + 0_{0\times 1}0_{1\times 0} \end{pmatrix}$$
$$= \begin{pmatrix} AB^T + uv^T & 0_{n\times 0} \\ 0_{0\times n} & 0_{0\times 0} \end{pmatrix} = AB^T + uv^T,$$

this rewrites as $(A \mid u)(B \mid v)^T = AB^T + uv^T$. This proves Proposition 7.265. □

We shall now give another, self-contained proof of Proposition 7.265, based upon the following simple fact:

Proposition 7.267. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times m}$ and $B \in \mathbb{K}^{n\times m}$. Then,

$$AB^T = \sum_{k=1}^{m} A_{\bullet,k} (B_{\bullet,k})^T .$$

(Here, we are using the notations introduced in Definition 6.128.)

Proof of Proposition 7.267. The matrix A is an $n \times m$-matrix (since $A \in \mathbb{K}^{n\times m}$). Write this $n \times m$-matrix A in the form $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$.

The matrix B is an $n \times m$-matrix (since $B \in \mathbb{K}^{n\times m}$). Write this $n \times m$-matrix B in the form $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$. From $B = (b_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$, we obtain $B^T = (b_{j,i})_{1\leq i\leq m,\ 1\leq j\leq n}$ (by the definition of B^T).

The definition of the product of two matrices yields

$$AB^T = \left(\sum_{k=1}^{m} a_{i,k} b_{j,k} \right)_{1\leq i\leq n,\ 1\leq j\leq n} \qquad (1497)$$

(since $A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$ and $B^T = (b_{j,i})_{1\leq i\leq m,\ 1\leq j\leq n}$).

Let us now fix $p \in \{1, 2, \ldots, m\}$. Then, $A_{\bullet,p}$ is the p-th column of the matrix A (by the definition of $A_{\bullet,p}$). Thus,

$$\begin{aligned} A_{\bullet,p} &= (\text{the } p\text{-th column of the matrix } A) \\ &= \begin{pmatrix} a_{1,p} \\ a_{2,p} \\ \vdots \\ a_{n,p} \end{pmatrix} \qquad \left(\text{since } A = (a_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}\right) \\ &= (a_{i,p})_{1\leq i\leq n,\ 1\leq j\leq 1} . \end{aligned}$$

The same argument (applied to B and $b_{i,j}$ instead of A and $a_{i,j}$) yields $B_{\bullet,p} = \left(b_{i,p}\right)_{1\leq i\leq n,\ 1\leq j\leq 1}$. Hence, $\left(B_{\bullet,p}\right)^T = \left(b_{j,p}\right)_{1\leq i\leq 1,\ 1\leq j\leq n}$ (by the definition of $\left(B_{\bullet,p}\right)^T$).

Now, the definition of the product of two matrices yields

$$A_{\bullet,p}\left(B_{\bullet,p}\right)^T = \left(\underbrace{\sum_{k=1}^{1} a_{i,p}b_{j,p}}_{=a_{i,p}b_{j,p}}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \qquad \left(\begin{array}{c}\text{since } A_{\bullet,p} = \left(a_{i,p}\right)_{1\leq i\leq n,\ 1\leq j\leq 1} \\ \text{and } \left(B_{\bullet,p}\right)^T = \left(b_{j,p}\right)_{1\leq i\leq 1,\ 1\leq j\leq n}\end{array}\right)$$
$$= \left(a_{i,p}b_{j,p}\right)_{1\leq i\leq n,\ 1\leq j\leq n}. \qquad (1498)$$

Now, let us forget that we fixed p. We thus have proven (1498) for each $p \in \{1, 2, \ldots, m\}$. Now,

$$\sum_{k=1}^{m} \underbrace{A_{\bullet,k}\left(B_{\bullet,k}\right)^T}_{\substack{=\left(a_{i,k}b_{j,k}\right)_{1\leq i\leq n,\ 1\leq j\leq n} \\ \text{(by (1498), applied to } p=k\text{)}}} = \sum_{k=1}^{m} \left(a_{i,k}b_{j,k}\right)_{1\leq i\leq n,\ 1\leq j\leq n} = \left(\sum_{k=1}^{m} a_{i,k}b_{j,k}\right)_{1\leq i\leq n,\ 1\leq j\leq n}.$$

Comparing this with (1497), we obtain $AB^T = \sum_{k=1}^{m} A_{\bullet,k}\left(B_{\bullet,k}\right)^T$. This proves Proposition 7.267. □

Second proof of Proposition 7.265. We have $(A \mid u) \in \mathbb{K}^{n\times(m+1)}$ (since $A \in \mathbb{K}^{n\times m}$ and $u \in \mathbb{K}^{n\times 1}$) and $(B \mid v) \in \mathbb{K}^{n\times(m+1)}$ (since $B \in \mathbb{K}^{n\times m}$ and $v \in \mathbb{K}^{n\times 1}$). Thus, Proposition 7.267 (applied to $m+1$, $(A \mid u)$ and $(B \mid v)$ instead of m, A and B) yields

$$\begin{aligned}
&(A \mid u)(B \mid v)^T \\
&= \sum_{k=1}^{m+1} (A \mid u)_{\bullet,k} \left((B \mid v)_{\bullet,k}\right)^T \\
&= \sum_{k=1}^{m} \underbrace{(A \mid u)_{\bullet,k}}_{\substack{=A_{\bullet,k} \\ \text{(by Proposition 6.134 (a),} \\ \text{applied to } u \text{ and } k \text{ instead of } v \text{ and } q)}} \left(\underbrace{(B \mid v)_{\bullet,k}}_{\substack{=B_{\bullet,k} \\ \text{(by Proposition 6.134 (a),} \\ \text{applied to } B \text{ and } k \text{ instead of } A \text{ and } q)}} \right)^T \\
&\qquad + \underbrace{(A \mid u)_{\bullet,m+1}}_{\substack{=u \\ \text{(by Proposition 6.134 (b),} \\ \text{applied to } u \text{ instead of } v)}} \left(\underbrace{(B \mid v)_{\bullet,m+1}}_{\substack{=v \\ \text{(by Proposition 6.134 (b),} \\ \text{applied to } B \text{ instead of } A)}} \right)^T \\
&\qquad \text{(here, we have split off the addend for } k = m+1 \text{ from the sum)} \\
&= \sum_{k=1}^{m} A_{\bullet,k} (B_{\bullet,k})^T + uv^T.
\end{aligned}$$

Comparing this with

$$\underbrace{AB^T}_{\substack{= \sum_{k=1}^{m} A_{\bullet,k}(B_{\bullet,k})^T \\ \text{(by Proposition 7.267)}}} + uv^T = \sum_{k=1}^{m} A_{\bullet,k} (B_{\bullet,k})^T + uv^T,$$

we obtain $(A \mid u)(B \mid v)^T = AB^T + uv^T$. Thus, Proposition 7.265 is proven again. □

Corollary 7.268. Let $n \in \mathbb{N}$. Let $u \in \mathbb{K}^{1\times n}$ and $v \in \mathbb{K}^{n\times 1}$. Let A be an $n \times n$-matrix. Then,

$$A + uv = (A \mid u) \left(I_n \mid v^T\right)^T.$$

(Here, we are using the notations introduced in Definition 6.132.)

Proof of Corollary 7.268. We have $v \in \mathbb{K}^{n\times 1}$ and thus $v^T \in \mathbb{K}^{1\times n}$. Also, $I_n \in \mathbb{K}^{n\times n}$. Thus, Proposition 7.265 (applied to n, I_n and v^T instead of m, B and v) yields

$$(A \mid u)\left(I_n \mid v^T\right)^T = A \underbrace{(I_n)^T}_{=I_n} + u \underbrace{\left(v^T\right)^T}_{\substack{=v \\ \text{(since } \left(C^T\right)^T = C \text{ for any} \\ \text{matrix } C)}} = \underbrace{AI_n}_{=A} + uv = A + uv.$$

This proves Corollary 7.268. □

Next, we state some simple facts about determinants:

Lemma 7.269. Let $n \in \mathbb{N}$. Then:

(a) Every $u \in \{1, 2, \ldots, n\}$ satisfies $\det\left((I_n)_{\sim u, \sim u}\right) = 1$.

(b) If u and v are two elements of $\{1, 2, \ldots, n\}$ such that $u \neq v$, then $\det\left((I_n)_{\sim u, \sim v}\right) = 0$.

Proof of Lemma 7.269. There are myriad ways to prove Lemma 7.269 (for example, one can prove part **(a)** by showing that $(I_n)_{\sim u, \sim u} = I_{n-1}$, and prove part **(b)** by arguing that the matrix $(I_n)_{\sim u, \sim v}$ has a row consisting of zeroes[669]). The proof we will now give is not the simplest one, but the shortest one (using what we have proven so far):

For any two objects i and j, define an element $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.
Then, $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of I_n).

Now, Theorem 6.100 (applied to $A = I_n$) yields $I_n \cdot \operatorname{adj}(I_n) = \operatorname{adj}(I_n) \cdot I_n = \det(I_n) \cdot I_n$. Thus, $\operatorname{adj}(I_n) \cdot I_n = \underbrace{\det(I_n)}_{=1} \cdot I_n = I_n$, so that $I_n = \operatorname{adj}(I_n) \cdot I_n = \operatorname{adj}(I_n)$ (since $BI_n = I_n$ for any $m \in \mathbb{N}$ and any $m \times n$-matrix B). Thus,

$$I_n = \operatorname{adj}(I_n) = \left((-1)^{i+j} \det\left((I_n)_{\sim j, \sim i}\right)\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$$

(by the definition of $\operatorname{adj}(I_n)$). Hence,

$$\left((-1)^{i+j} \det\left((I_n)_{\sim j, \sim i}\right)\right)_{1 \leq i \leq n,\ 1 \leq j \leq n} = I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}.$$

In other words,

$$(-1)^{i+j} \det\left((I_n)_{\sim j, \sim i}\right) = \delta_{i,j} \qquad (1499)$$

for every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$.

(a) Let $u \in \{1, 2, \ldots, n\}$. Then, (1499) (applied to $i = u$ and $j = u$) yields

$$(-1)^{u+u} \det\left((I_n)_{\sim u, \sim u}\right) = \delta_{u,u} = 1 \qquad (\text{since } u = u).$$

Comparing this with $\underbrace{(-1)^{u+u}}_{\substack{=1 \\ (\text{since } u+u=2u \text{ is even})}} \det\left((I_n)_{\sim u, \sim u}\right) = \det\left((I_n)_{\sim u, \sim u}\right)$, we obtain $\det\left((I_n)_{\sim u, \sim u}\right) = 1$. This proves Lemma 7.269 **(a)**.

[669] or, what also suffices, a column consisting of zeroes

(b) Let u and v be two elements of $\{1, 2, \ldots, n\}$ such that $u \neq v$. Then, (1499) (applied to $i = v$ and $j = u$) yields

$$(-1)^{v+u} \det \left((I_n)_{\sim u, \sim v} \right) = \delta_{v,u} = 0 \qquad (\text{since } v \neq u \text{ (since } u \neq v)) .$$

Multiplying both sides of this equality by $(-1)^{v+u}$, we obtain

$$(-1)^{v+u} (-1)^{v+u} \det \left((I_n)_{\sim u, \sim v} \right) = 0.$$

Comparing this with

$$\underbrace{(-1)^{v+u} (-1)^{v+u}}_{\substack{= (-1)^{(v+u)+(v+u)} = 1 \\ (\text{since } (v+u)+(v+u) = 2(v+u) \text{ is even})}} \det \left((I_n)_{\sim u, \sim v} \right) = \det \left((I_n)_{\sim u, \sim v} \right),$$

we obtain $\det \left((I_n)_{\sim u, \sim v} \right) = 0$. Lemma 7.269 **(b)** is thus proven. □

Proposition 7.270. Let $n \in \mathbb{N}$. Let $v = (v_1, v_2, \ldots, v_n)^T \in \mathbb{K}^{n \times 1}$ and $k \in \{1, 2, \ldots, n\}$. Then,

$$\det \left((I_n \mid v)_{\bullet, \sim k} \right) = (-1)^{n+k} v_k.$$

(Here, we are using the notations introduced in Definition 6.128.)

Proof of Proposition 7.270. We have $I_n \in \mathbb{K}^{n \times n}$ and thus $(I_n)_{\bullet, \sim k} \in \mathbb{K}^{n \times (n-1)}$. Every $i \in \{1, 2, \ldots, n\}$ satisfies

$$\left((I_n)_{\bullet, \sim k} \right)_{\sim i, \bullet} = (I_n)_{\sim i, \sim k} \tag{1500}$$

[670].

Proposition 6.134 **(d)** (applied to $m = n$, $A = I_n$ and $q = k$) yields $(I_n \mid v)_{\bullet, \sim k} =$

[670] *Proof of (1500):* Let $i \in \{1, 2, \ldots, n\}$. Proposition 6.130 **(c)** (applied to n, I_n, i and k instead of m, A, u and v) shows that $\left((I_n)_{\bullet, \sim k} \right)_{\sim i, \bullet} = \left((I_n)_{\sim i, \bullet} \right)_{\bullet, \sim k} = (I_n)_{\sim i, \sim k}$. This proves (1500).

$\left((I_n)_{\bullet,\sim k} \mid v \right)$. Hence,

$$\begin{aligned}
&\det \underbrace{\left((I_n \mid v)_{\bullet,\sim k} \right)}_{=\left((I_n)_{\bullet,\sim k} \mid v \right)} \\
&= \det \left((I_n)_{\bullet,\sim k} \mid v \right) \\
&= \underbrace{\sum_{i=1}^{n}}_{=\sum_{i\in\{1,2,\ldots,n\}}} (-1)^{n+i} v_i \det \left(\underbrace{\left((I_n)_{\bullet,\sim k} \right)_{\sim i,\bullet}}_{\substack{=(I_n)_{\sim i,\sim k} \\ \text{(by (1500))}}} \right) \\
&\qquad \left(\text{by Proposition 6.135 } \textbf{(a)}\text{, applied to } A = (I_n)_{\bullet,\sim k} \right) \\
&= \sum_{i\in\{1,2,\ldots,n\}} (-1)^{n+i} v_i \det \left((I_n)_{\sim i,\sim k} \right) \\
&= (-1)^{n+k} v_k \underbrace{\det \left((I_n)_{\sim k,\sim k} \right)}_{\substack{=1 \\ \text{(by Lemma 7.269 } \textbf{(a)}\text{,} \\ \text{applied to } k \text{ instead of } u\text{)}}} + \sum_{\substack{i\in\{1,2,\ldots,n\}; \\ i\neq k}} (-1)^{n+i} v_i \underbrace{\det \left((I_n)_{\sim i,\sim k} \right)}_{\substack{=0 \\ \text{(by Lemma 7.269 } \textbf{(b)}\text{,} \\ \text{applied to } i \text{ and } k \\ \text{instead of } u \text{ and } v\text{)}}} \\
&\qquad \left(\text{here, we have split off the addend for } i = k \text{ from the sum} \right) \\
&= (-1)^{n+k} v_k + \underbrace{\sum_{\substack{i\in\{1,2,\ldots,n\}; \\ i\neq k}} (-1)^{n+i} v_i 0}_{=0} = (-1)^{n+k} v_k.
\end{aligned}$$

This proves Proposition 7.270. ☐

Proposition 7.271. Let $n \in \mathbb{N}$. Let $A \in \mathbb{K}^{n\times(n+1)}$ and $B \in \mathbb{K}^{n\times(n+1)}$. Then,

$$\det \left(AB^T \right) = \sum_{k=1}^{n+1} \det \left(A_{\bullet,\sim k} \right) \det \left(B_{\bullet,\sim k} \right).$$

(Here, we are using the notations introduced in Definition 6.128.)

Proof of Proposition 7.271. Let us use the notations introduced in Definition 6.31. Clearly, $n+1$ is a positive integer (since $n \in \mathbb{N}$).

Also, A is an $n \times (n+1)$-matrix (since $A \in \mathbb{K}^{n\times(n+1)}$). In other words, A is an $((n+1)-1) \times (n+1)$-matrix (since $n = (n+1)-1$).

Also, B is an $n \times (n+1)$-matrix (since $B \in \mathbb{K}^{n\times(n+1)}$). Hence, B^T is an $(n+1) \times n$-matrix. In other words, B^T is an $(n+1) \times ((n+1)-1)$-matrix (since $n = (n+1) - 1$).

Now, every $k \in \{1, 2, \ldots, n+1\}$ satisfies

$$\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n+1} A = A_{\bullet,\sim k} \tag{1501}$$

[671] and

$$\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1} \left(B^T\right) = \left(B_{\bullet,\sim k}\right)^T \tag{1502}$$

[672] and

$$\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1} \left(B^T\right)\right) = \det\left(B_{\bullet,\sim k}\right) \tag{1503}$$

[673].

Now, Lemma 7.180 (applied to $n+1$ and B^T instead of n and B) yields

$$\begin{aligned}\det\left(AB^T\right) &= \sum_{k=1}^{n+1} \det\left(\underbrace{\operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n+1} A}_{\substack{=A_{\bullet,\sim k} \\ \text{(by (1501))}}}\right) \cdot \underbrace{\det\left(\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1}\left(B^T\right)\right)}_{\substack{=\det\left(B_{\bullet,\sim k}\right) \\ \text{(by (1503))}}} \\ &= \sum_{k=1}^{n+1} \det\left(A_{\bullet,\sim k}\right) \det\left(B_{\bullet,\sim k}\right).\end{aligned}$$

This proves Proposition 7.271. □

The next lemma is an easy consequence of results proven before:

---

[671] *Proof of (1501):* Let $k \in \{1, 2, \ldots, n+1\}$. Then, the definition of $A_{\bullet,\sim k}$ yields $A_{\bullet,\sim k} = \operatorname{cols}_{1,2,\ldots,\widehat{k},\ldots,n+1} A$ (since $A \in \mathbb{K}^{n\times(n+1)}$). This proves (1501).

[672] *Proof of (1502):* Let $k \in \{1, 2, \ldots, n+1\}$. Then, the definition of $\left(B^T\right)_{\sim k,\bullet}$ yields $\left(B^T\right)_{\sim k,\bullet} = \operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1}\left(B^T\right)$ (since $B^T \in \mathbb{K}^{(n+1)\times n}$ (since B^T is an $(n+1)\times n$-matrix)). But Lemma 6.148 (applied to $m = n+1$ and $r = k$) yields $\left(B^T\right)_{\sim k,\bullet} = \left(B_{\bullet,\sim k}\right)^T$. Comparing this with $\left(B^T\right)_{\sim k,\bullet} = \operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1}\left(B^T\right)$, we obtain $\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1}\left(B^T\right) = \left(B_{\bullet,\sim k}\right)^T$. This proves (1502).

[673] *Proof of (1503):* Let $k \in \{1, 2, \ldots, n+1\}$. From $B \in \mathbb{K}^{n\times(n+1)}$, we obtain $B_{\bullet,\sim k} \in \mathbb{K}^{n\times n}$. In other words, $B_{\bullet,\sim k}$ is an $n\times n$-matrix. Hence, Exercise 6.4 (applied to $B_{\bullet,\sim k}$ instead of A) yields $\det\left(\left(B_{\bullet,\sim k}\right)^T\right) = \det\left(B_{\bullet,\sim k}\right)$. Now,

$$\det\left(\underbrace{\operatorname{rows}_{1,2,\ldots,\widehat{k},\ldots,n+1}\left(B^T\right)}_{\substack{=\left(B_{\bullet,\sim k}\right)^T \\ \text{(by (1502))}}}\right) = \det\left(\left(B_{\bullet,\sim k}\right)^T\right) = \det\left(B_{\bullet,\sim k}\right).$$

This proves (1503).

Lemma 7.272. Let $n \in \mathbb{N}$. Let $u = (u_1, u_2, \ldots, u_n)^T \in \mathbb{K}^{n\times 1}$ be a column vector with n entries. Let A be an $n \times n$-matrix. Let $k \in \{1, 2, \ldots, n\}$. Then,

$$\det\left((A \mid u)_{\bullet,\sim k}\right) = \sum_{j=1}^{n} (-1)^{n+j} u_j \det\left(A_{\sim j,\sim k}\right).$$

Proof of Lemma 7.272. From $A \in \mathbb{K}^{n\times n}$ and $u \in \mathbb{K}^{n\times 1}$, we obtain $(A \mid u) \in \mathbb{K}^{n\times(n+1)}$.

From $k \in \{1, 2, \ldots, n\}$, we obtain $A_{\bullet,\sim k} \in \mathbb{K}^{n\times(n-1)}$ (since $A \in \mathbb{K}^{n\times n}$). Proposition 6.134 **(c)** (applied to n, u and k instead of m, v and q) shows that $(A \mid u)_{\bullet,\sim k} = (A_{\bullet,\sim k} \mid u)$. Hence,

$$\begin{aligned}
\det\Big(\underbrace{(A \mid u)_{\bullet,\sim k}}_{=(A_{\bullet,\sim k}\mid u)}\Big) &= \det\left(A_{\bullet,\sim k} \mid u\right) \\
&= \sum_{i=1}^{n} (-1)^{n+i} u_i \det\Bigg(\underbrace{\left(A_{\bullet,\sim k}\right)_{\sim i,\bullet}}_{\substack{=A_{\sim i,\sim k} \\ \text{(since Proposition 6.130 (c) (applied to} \\ n,\ i \text{ and } k \text{ instead of } m,\ u \text{ and } v\text{) shows that} \\ (A_{\bullet,\sim k})_{\sim i,\bullet}=(A_{\sim i,\bullet})_{\bullet,\sim k}=A_{\sim i,\sim k})}}\Bigg) \\
&\qquad \left(\begin{array}{c}\text{by Proposition 6.135 (a), applied to} \\ A_{\bullet,\sim k},\ u \text{ and } u_i \text{ instead of } A,\ v \text{ and } v_i\end{array}\right) \\
&= \sum_{i=1}^{n} (-1)^{n+i} u_i \det\left(A_{\sim i,\sim k}\right) = \sum_{j=1}^{n} (-1)^{n+j} u_j \det\left(A_{\sim j,\sim k}\right)
\end{aligned}$$

(here, we have renamed the summation index i as j). This proves Lemma 7.272. □

Now we are more than ready to easily prove Theorem 7.262:

Proof of Theorem 7.262. We shall use the notations introduced in Definition 6.81, in Definition 6.132 and in Definition 6.128.

We know that u is a column vector with n entries; in other words, $u \in \mathbb{K}^{n\times 1}$. Also, v is a row vector with n entries; in other words, $v \in \mathbb{K}^{1\times n}$. Hence, $v^T \in \mathbb{K}^{n\times 1}$.

From $A \in \mathbb{K}^{n\times n}$ and $u \in \mathbb{K}^{n\times 1}$, we obtain $(A \mid u) \in \mathbb{K}^{n\times(n+1)}$. From $I_n \in \mathbb{K}^{n\times n}$ and $v^T \in \mathbb{K}^{n\times 1}$, we obtain $\left(I_n \mid v^T\right) \in \mathbb{K}^{n\times(n+1)}$.

Write the vector v in the form $v = (v_1, v_2, \ldots, v_n)$. (This is possible, since v is a row vector with n entries.) From $v = (v_1, v_2, \ldots, v_n)$, we obtain $v^T = (v_1, v_2, \ldots, v_n)^T$.

Write the vector u in the form $u = (u_1, u_2, \ldots, u_n)^T$. (This is possible, since u is a column vector with n entries.)

Now,

$$\det \left(\underbrace{A + uv}_{\substack{=(A|u)\left(I_n|v^T\right)^T \\ \text{(by Corollary 7.268)}}} \right)$$

$$= \det \left((A \mid u) \left(I_n \mid v^T \right)^T \right) = \sum_{k=1}^{n+1} \det \left((A \mid u)_{\bullet, \sim k} \right) \det \left(\left(I_n \mid v^T \right)_{\bullet, \sim k} \right)$$

$$\left(\begin{array}{c} \text{by Proposition 7.271, applied to } (A \mid u) \text{ and } \left(I_n \mid v^T\right) \\ \text{instead of } A \text{ and } B \end{array} \right)$$

$$= \sum_{k=1}^{n} \underbrace{\det \left((A \mid u)_{\bullet, \sim k} \right)}_{\substack{= \sum_{j=1}^{n} (-1)^{n+j} u_j \det\left(A_{\sim j, \sim k}\right) \\ \text{(by Lemma 7.272)}}} \underbrace{\det \left(\left(I_n \mid v^T \right)_{\bullet, \sim k} \right)}_{\substack{=(-1)^{n+k} v_k \\ \text{(by Proposition 7.270,} \\ \text{applied to } v^T \text{ instead of } v)}}$$

$$+ \det \left(\underbrace{(A \mid u)_{\bullet, \sim (n+1)}}_{\substack{=A \\ \text{(by Proposition 6.134 \textbf{(d)}, applied} \\ \text{to } n \text{ and } u \text{ instead of } m \text{ and } v)}} \right) \det \left(\underbrace{\left(I_n \mid v^T \right)_{\bullet, \sim (n+1)}}_{\substack{=I_n \\ \text{(by Proposition 6.134 \textbf{(d)}, applied} \\ \text{to } n,\, I_n \text{ and } v^T \text{ instead of } m,\, A \text{ and } v)}} \right)$$

(here, we have split off the addend for $k = n + 1$ from the sum)

$$= \underbrace{\sum_{k=1}^{n} \left(\sum_{j=1}^{n} (-1)^{n+j} u_j \det \left(A_{\sim j, \sim k} \right) \right) \cdot (-1)^{n+k} v_k}_{= \sum_{k=1}^{n} \sum_{j=1}^{n} (-1)^{n+j} u_j \det\left(A_{\sim j, \sim k}\right) \cdot (-1)^{n+k} v_k} + \det A \cdot \underbrace{\det \left(I_n \right)}_{=1}$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{n} (-1)^{n+j} u_j \det \left(A_{\sim j, \sim k} \right) \cdot (-1)^{n+k} v_k + \det A.$$

Subtracting $\det A$ from both sides of this equality, we obtain

$$\begin{aligned}
&\det(A+uv) - \det A \\
&= \sum_{k=1}^{n} \sum_{j=1}^{n} (-1)^{n+j} u_j \det\left(A_{\sim j,\sim k}\right) \cdot (-1)^{n+k} v_k \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \underbrace{(-1)^{n+j} u_j \det\left(A_{\sim j,\sim i}\right) \cdot (-1)^{n+i} v_i}_{=(-1)^{n+i}(-1)^{n+j} u_j v_i \det\left(A_{\sim j,\sim i}\right)} \\
&\qquad \left(\text{here, we have renamed the summation index } k \text{ as } i \text{ in the first sum}\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \underbrace{(-1)^{n+i} (-1)^{n+j}}_{\substack{=(-1)^{(n+i)+(n+j)}=(-1)^{i+j} \\ (\text{since } (n+i)+(n+j)=2n+i+j\equiv i+j \bmod 2)}} u_j v_i \det\left(A_{\sim j,\sim i}\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right).
\end{aligned}$$

Comparing this with

$$\operatorname{ent}(v(\operatorname{adj} A) u) = \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right) \qquad (\text{by Corollary 7.264}),$$

we obtain $\det(A+uv) - \det A = \operatorname{ent}(v(\operatorname{adj} A) u)$. In other words, $\det(A+uv) = \det A + \operatorname{ent}(v(\operatorname{adj} A) u)$. This proves Theorem 7.262. □

Now, Theorem 7.262 is proven; in other words, Exercise 6.59 is solved (since Theorem 7.262 is just a restatement of the claim of this exercise).

7.127. Solution to Exercise 6.60

In this section, we shall use the notations introduced in Definition 6.31, in Definition 6.81, in Definition 6.89, in Definition 6.128 and in Definition 7.261.

We begin with some simple lemmas.

Lemma 7.273. Let $n \in \mathbb{N}$. Let $u \in \mathbb{K}^{n\times 1}$ be a column vector with n entries, and let $v = (v_1, v_2, \ldots, v_n) \in \mathbb{K}^{1\times n}$ be a row vector with n entries. (Thus, uv is an $n \times n$-matrix, whereas vu is a 1×1-matrix.) Let $h \in \mathbb{K}$. Let H be the 1×1-matrix $\begin{pmatrix} h \end{pmatrix} \in \mathbb{K}^{1\times 1}$. Let $A \in \mathbb{K}^{n\times n}$ be an $n \times n$-matrix. Let C be the $(n+1) \times (n+1)$-matrix $\begin{pmatrix} A & u \\ v & H \end{pmatrix}$. Write this $(n+1) \times (n+1)$-matrix C in the form $C = \left(c_{i,j}\right)_{1\leq i\leq n+1,\ 1\leq j\leq n+1}$. Then:

(a) We have $c_{n+1,q} = v_q$ for each $q \in \{1, 2, \ldots, n\}$.

(b) We have $c_{n+1,n+1} = h$.

(c) We have $C_{\sim(n+1),\bullet} = (A \mid u)$.
(d) We have $C_{\sim(n+1),\sim q} = (A \mid u)_{\bullet,\sim q}$ for each $q \in \{1, 2, \ldots, n+1\}$.
(e) We have $C_{\sim(n+1),\sim(n+1)} = A$.

Proof of Lemma 7.273. All parts of Lemma 7.273 are obvious from a look at the matrix $C = \begin{pmatrix} A & u \\ v & H \end{pmatrix}$: For example, part **(d)** says that removing the $(n+1)$-st row and the q-th column (for some $q \in \{1, 2, \ldots, n+1\}$) from this matrix yields the same result as attaching the column vector u to A at its right and then removing the q-th column (which should be obvious). Turning this into a rigorous proof is a straightforward exercise in bookkeeping (using (435), Proposition 6.130 and Proposition 6.134). □

Lemma 7.274. Let $n \in \mathbb{N}$. Let $u = (u_1, u_2, \ldots, u_n)^T \in \mathbb{K}^{n \times 1}$ be a column vector with n entries, and let $v = (v_1, v_2, \ldots, v_n) \in \mathbb{K}^{1 \times n}$ be a row vector with n entries. Let $h \in \mathbb{K}$. Let H be the 1×1-matrix $\begin{pmatrix} h \end{pmatrix} \in \mathbb{K}^{1 \times 1}$. Let $A \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix. Then,

$$\det \begin{pmatrix} A & u \\ v & H \end{pmatrix} = h \det A - \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det \left(A_{\sim j, \sim i} \right).$$

Proof of Lemma 7.274. Let C be the $(n+1) \times (n+1)$-matrix $\begin{pmatrix} A & u \\ v & H \end{pmatrix}$. Write this $(n+1) \times (n+1)$-matrix C in the form $C = \left(c_{i,j} \right)_{1 \leq i \leq n+1,\ 1 \leq j \leq n+1}$.

Thus, Theorem 6.82 **(a)** (applied to $n+1$, C, $c_{i,j}$ and $n+1$ instead of n, A, $a_{i,j}$ and

$n+1$) yields

$$\begin{aligned}
&\det C \\
&= \sum_{q=1}^{n+1} (-1)^{(n+1)+q} c_{n+1,q} \det \left(C_{\sim(n+1),\sim q} \right) \\
&\qquad \left(\text{since } C = \left(c_{i,j} \right)_{1 \leq i \leq n+1,\ 1 \leq j \leq n+1} \right) \\
&= \sum_{q=1}^{n} (-1)^{(n+1)+q} \underbrace{c_{n+1,q}}_{\substack{=v_q \\ \text{(by Lemma 7.273 \textbf{(a)})}}} \det \left(\underbrace{C_{\sim(n+1),\sim q}}_{\substack{=(A|u)_{\bullet,\sim q} \\ \text{(by Lemma 7.273 \textbf{(d)}} \\ \text{(since } q\in\{1,2,\ldots,n\}\subseteq\{1,2,\ldots,n+1\}))}} \right) \\
&\qquad + \underbrace{(-1)^{(n+1)+(n+1)}}_{\substack{=1 \\ \text{(since } (n+1)+(n+1)=2(n+1) \text{ is even)}}} \underbrace{c_{n+1,n+1}}_{\substack{=h \\ \text{(by Lemma 7.273 \textbf{(b)})}}} \det \left(\underbrace{C_{\sim(n+1),\sim(n+1)}}_{\substack{=A \\ \text{(by Lemma 7.273 \textbf{(e)})}}} \right) \\
&\qquad \left(\text{here, we have split off the addend for } q = n+1 \text{ from the sum} \right) \\
&= \sum_{q=1}^{n} (-1)^{(n+1)+q} v_q \det \left((A \mid u)_{\bullet,\sim q} \right) + h \det A.
\end{aligned}$$

Subtracting $h \det A$ from both sides of this equality, we obtain

$$\begin{aligned}
&\det C - h\det A \\
&= \sum_{q=1}^{n} (-1)^{(n+1)+q} v_q \underbrace{\det\left((A \mid u)_{\bullet,\sim q}\right)}_{\substack{=\sum_{j=1}^{n}(-1)^{n+j}u_j\det\left(A_{\sim j,\sim q}\right) \\ \text{(by Lemma 7.272 (applied to } k=q\text{))}}} \\
&= \sum_{q=1}^{n} (-1)^{(n+1)+q} v_q \left(\sum_{j=1}^{n} (-1)^{n+j} u_j \det\left(A_{\sim j,\sim q}\right)\right) \\
&= \sum_{q=1}^{n}\sum_{j=1}^{n} (-1)^{(n+1)+q} v_q (-1)^{n+j} u_j \det\left(A_{\sim j,\sim q}\right) \\
&= \sum_{q=1}^{n}\sum_{j=1}^{n} \underbrace{(-1)^{(n+1)+q}(-1)^{n+j}}_{\substack{=(-1)^{((n+1)+q)+(n+j)}=(-1)^{q+j+1} \\ \text{(since } ((n+1)+q)+(n+j)=(q+j+1)+2n\equiv q+j+1 \bmod 2\text{)}}} u_j v_q \det\left(A_{\sim j,\sim q}\right) \\
&= \sum_{q=1}^{n}\sum_{j=1}^{n} \underbrace{(-1)^{q+j+1}}_{=-(-1)^{q+j}} u_j v_q \det\left(A_{\sim j,\sim q}\right) \\
&= -\sum_{q=1}^{n}\sum_{j=1}^{n} (-1)^{q+j} u_j v_q \det\left(A_{\sim j,\sim q}\right) = -\sum_{i=1}^{n}\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right)
\end{aligned}$$

(here, we have renamed the summation index q as i).

Adding $h \det A$ to both sides of this equality, we obtain

$$\det C = h\det A - \sum_{i=1}^{n}\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right).$$

In view of $C = \begin{pmatrix} A & u \\ v & H \end{pmatrix}$ (by the definition of C), this rewrites as

$$\det\begin{pmatrix} A & u \\ v & H \end{pmatrix} = h\det A - \sum_{i=1}^{n}\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j,\sim i}\right).$$

This proves Lemma 7.274. □

We can now easily obtain the following corollary, which is just a restatement of Exercise 6.60 **(a)**:

Corollary 7.275. Let $n \in \mathbb{N}$. Let $u \in \mathbb{K}^{n\times 1}$ be a column vector with n entries, and let $v \in \mathbb{K}^{1\times n}$ be a row vector with n entries. (Thus, uv is an $n \times n$-matrix, whereas vu is a 1×1-matrix.) Let $h \in \mathbb{K}$. Let H be the 1×1-matrix $\begin{pmatrix} h \end{pmatrix} \in \mathbb{K}^{1 \times 1}$. Let $A \in \mathbb{K}^{n \times n}$ be an $n \times n$-matrix. Then,

$$\det \begin{pmatrix} A & u \\ v & H \end{pmatrix} = h \det A - \operatorname{ent}\left(v \left(\operatorname{adj} A\right) u\right).$$

(Here, we are using the notation from Definition 7.261.)

Proof of Corollary 7.275. Write the column vector $u \in \mathbb{K}^{n \times 1}$ in the form $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$.

Thus, $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \left(u_1, u_2, \ldots, u_n\right)^T$.

Write the row vector $v \in \mathbb{K}^{1 \times n}$ in the form $v = \left(v_1, v_2, \ldots, v_n\right)$. Thus, Corollary 7.264 yields

$$\operatorname{ent}\left(v \left(\operatorname{adj} A\right) u\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j, \sim i}\right). \qquad (1504)$$

But Lemma 7.274 yields

$$\det \begin{pmatrix} A & u \\ v & H \end{pmatrix} = h \det A - \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(A_{\sim j, \sim i}\right)}_{\substack{= \operatorname{ent}(v(\operatorname{adj} A)u) \\ \text{(by (1504))}}}$$
$$= h \det A - \operatorname{ent}\left(v \left(\operatorname{adj} A\right) u\right).$$

This proves Corollary 7.275. ☐

We now take aim at part **(b)** of Exercise 6.60. We begin with a lemma that computes the $(n-1) \times (n-1)$-minors of a diagonal matrix:

Lemma 7.276. Let $n \in \mathbb{N}$. For every two objects i and j, define $\delta_{i,j} \in \mathbb{K}$ by $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$.

Let $d_1, d_2, \ldots, d_n$ be n elements of $\mathbb{K}$. Let D be the $n \times n$-matrix $\left(d_i \delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Let $p \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, n\}$. Then,

$$\det\left(D_{\sim p, \sim q}\right) = \delta_{p,q} \prod_{\substack{j \in \{1, 2, \ldots, n\}; \\ j \neq p}} d_j.$$

Proof of Lemma 7.276. We shall use the notations from Definition 6.153.

Define a subset P of $\{1,2,\ldots,n\}$ by $P = \{1,2,\ldots,n\} \setminus \{p\}$.

Define a subset Q of $\{1,2,\ldots,n\}$ by $Q = \{1,2,\ldots,n\} \setminus \{q\}$.

Clearly, $P = Q$ holds if and only if $p = q$ holds. Thus, $\delta_{P,Q} = \delta_{p,q}$. Also, clearly, $|P| = |Q| = n-1$.

The definition of $D_{\sim p,\sim q}$ yields

$$D_{\sim p,\sim q} = \operatorname{sub}_{1,2,\ldots,\widehat{p},\ldots,n}^{1,2,\ldots,\widehat{q},\ldots,n} D. \tag{1505}$$

The definition of the list $(1,2,\ldots,\widehat{p},\ldots,n)$ yields

$$(1,2,\ldots,\widehat{p},\ldots,n) = (1,2,\ldots,p-1,p+1,p+2,\ldots,n).$$

Meanwhile, the definition of $w(P)$ shows that $w(P)$ is the list of all elements of P in increasing order (with no repetitions); thus, $w(P) = (1,2,\ldots,p-1,p+1,p+2,\ldots,n)$ (because $P = \{1,2,\ldots,n\} \setminus \{p\} = \{1,2,\ldots,p-1,p+1,p+2,\ldots,n\}$). Comparing these two equalities, we obtain $w(P) = (1,2,\ldots,\widehat{p},\ldots,n)$. Similarly, $w(Q) = (1,2,\ldots,\widehat{q},\ldots,n)$. Using the preceding two equalities, we find

$$\operatorname{sub}_{w(P)}^{w(Q)} D = \operatorname{sub}_{(1,2,\ldots,\widehat{p},\ldots,n)}^{(1,2,\ldots,\widehat{q},\ldots,n)} D = \operatorname{sub}_{1,2,\ldots,\widehat{p},\ldots,n}^{1,2,\ldots,\widehat{q},\ldots,n} D.$$

Comparing this with (1505), we find $D_{\sim p,\sim q} = \operatorname{sub}_{w(P)}^{w(Q)} D$. Hence,

$$\begin{aligned}
\det\left(D_{\sim p,\sim q}\right) &= \det\left(\operatorname{sub}_{w(P)}^{w(Q)} D\right) = \underbrace{\delta_{P,Q}}_{=\delta_{p,q}} \underbrace{\prod_{i\in P}}_{\substack{=\prod_{i\in\{1,2,\ldots,n\}\setminus\{p\}} \\ (\text{since } P=\{1,2,\ldots,n\}\setminus\{p\})}} d_i \\
&\qquad (\text{by Lemma 6.163}) \\
&= \delta_{p,q} \prod_{i\in\{1,2,\ldots,n\}\setminus\{p\}} d_i = \delta_{p,q} \underbrace{\prod_{j\in\{1,2,\ldots,n\}\setminus\{p\}}}_{=\prod_{j\in\{1,2,\ldots,n\};\ j\neq p}} d_j \\
&\qquad \left(\begin{array}{c}\text{here, we have renamed the index } i \text{ as } j \\ \text{in the product}\end{array}\right) \\
&= \delta_{p,q} \prod_{\substack{j\in\{1,2,\ldots,n\};\\ j\neq p}} d_j.
\end{aligned}$$

This proves Lemma 7.276. □

Notice that Lemma 7.276 generalizes Lemma 7.269 (because if we set $d_i = 1$ in Lemma 7.276, then the resulting matrix D is I_n).

We are now finally able to solve Exercise 6.60:

Solution to Exercise 6.60. **(a)** Let $A \in \mathbb{K}^{n\times n}$ be an $n \times n$-matrix. Then, Corollary 7.275 yields

$$\det \begin{pmatrix} A & u \\ v & H \end{pmatrix} = h \det A - \operatorname{ent}\left(v\left(\operatorname{adj} A\right) u\right) \tag{1506}$$

(where we are using the notation from Definition 7.261). But if we regard the 1×1-matrix $v\left(\operatorname{adj} A\right) u$ as an element of $\mathbb{K}$, then $\operatorname{ent}\left(v\left(\operatorname{adj} A\right) u\right) = v\left(\operatorname{adj} A\right) u$, and therefore (1506) simplifies as follows:

$$\det \begin{pmatrix} A & u \\ v & H \end{pmatrix} = h \det A - \underbrace{\operatorname{ent}\left(v\left(\operatorname{adj} A\right) u\right)}_{=v(\operatorname{adj} A)u} = h \det A - v\left(\operatorname{adj} A\right) u.$$

This solves Exercise 6.60 **(a)**.

(b) We have $d_i \delta_{i,j} = 0$ for every $(i,j) \in \{1,2,\ldots,n\}^2$ satisfying $i < j$ [674]. Hence, Exercise 6.3 (applied to D and $d_i \delta_{i,j}$ instead of A and $a_{i,j}$) yields

$$\begin{aligned} \det D &= \left(d_1 \delta_{1,1}\right)\left(d_2 \delta_{2,2}\right) \cdots \left(d_n \delta_{n,n}\right) \qquad \left(\text{since } D = \left(d_i \delta_{i,j}\right)_{1\le i\le n,\ 1\le j\le n}\right) \\ &= \prod_{k=1}^{n} \left(d_k \underbrace{\delta_{k,k}}_{\substack{=1 \\ (\text{since } k=k)}}\right) = \prod_{k=1}^{n} d_k = d_1 d_2 \cdots d_n. \end{aligned}$$

Applying Lemma 7.274 to $A = D$, we obtain

$$\det \begin{pmatrix} D & u \\ v & H \end{pmatrix} = h \det D - \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(D_{\sim j,\sim i}\right). \tag{1507}$$

[674] *Proof.* Let $(i,j) \in \{1,2,\ldots,n\}^2$ be such that $i < j$. From $i < j$, we obtain $i \neq j$; thus, $\delta_{i,j} = 0$. Hence, $d_i \underbrace{\delta_{i,j}}_{=0} = 0$, qed.

But each $i \in \{1, 2, \ldots, n\}$ satisfies

$$\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(D_{\sim j, \sim i}\right)$$
$$= \underbrace{\sum_{p=1}^{n}}_{= \sum_{p \in \{1,2,\ldots,n\}}} (-1)^{i+p} u_p v_i \underbrace{\det\left(D_{\sim p, \sim i}\right)}_{\substack{= \delta_{p,i} \prod_{j \in \{1,2,\ldots,n\};\ j \neq p} d_j \\ \text{(by Lemma 7.276 (applied to } q=i\text{))}}}$$
$$\left(\text{here, we have renamed the summation index } j \text{ as } p\right)$$
$$= \sum_{p \in \{1,2,\ldots,n\}} (-1)^{i+p} u_p v_i \delta_{p,i} \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq p}} d_j$$
$$= \underbrace{(-1)^{i+i}}_{\substack{=1 \\ \text{(since } i+i=2i \text{ is even)}}} u_i v_i \underbrace{\delta_{i,i}}_{\substack{=1 \\ \text{(since } i=i\text{)}}} \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq i}} d_j + \sum_{\substack{p \in \{1,2,\ldots,n\};\\ p \neq i}} (-1)^{i+p} u_p v_i \underbrace{\delta_{p,i}}_{\substack{=0 \\ \text{(since } p\neq i\text{)}}} \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq p}} d_j$$
$$\left(\begin{array}{c}\text{here, we have split off the addend for } p = i \text{ from the sum,} \\ \text{since } i \in \{1, 2, \ldots, n\}\end{array}\right)$$
$$= u_i v_i \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq i}} d_j + \underbrace{\sum_{\substack{p \in \{1,2,\ldots,n\};\\ p \neq i}} (-1)^{i+p} u_p v_i 0 \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq p}} d_j}_{=0}$$
$$= u_i v_i \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq i}} d_j. \qquad (1508)$$

Hence, (1507) becomes

$$\det \begin{pmatrix} D & u \\ v & H \end{pmatrix} = h \underbrace{\det D}_{= d_1 d_2 \cdots d_n} - \sum_{i=1}^{n} \underbrace{\sum_{j=1}^{n} (-1)^{i+j} u_j v_i \det\left(D_{\sim j, \sim i}\right)}_{\substack{= u_i v_i \prod_{j \in \{1,2,\ldots,n\};\ j \neq i} d_j \\ \text{(by (1508))}}}$$
$$= h \cdot (d_1 d_2 \cdots d_n) - \sum_{i=1}^{n} u_i v_i \prod_{\substack{j \in \{1,2,\ldots,n\};\\ j \neq i}} d_j.$$

This solves Exercise 6.60 **(b)**. □

8. Appendix: Old citations

Some places (articles, internet postings, etc.) cite older versions of these notes. Most of the time, the numbering of theorems, propositions, formulas and even chapters has shifted since these older versions were written; thus, these citations no longer point to the correct location of the results they want to cite. The following tables list the results cited in all citations known to me, and their locations in the current version of these notes.

item cited	its current location
Additional exercise 23 in the version of 19 May 2016	Exercise 6.59
Corollary 5.44 in the version of 29 May 2016	Corollary 6.45
Corollary 5.85 in the version of 29 May 2016	Corollary 6.102
Exercise 41 in the version of 4 September 2016	Exercise 6.22
Lemma 5.49 in the version of 2016-09-04	Lemma 6.49
Exercise 36 in the version of 6 October 2016	Exercise 6.17
Proposition 2.25 in the version of 2016-12-22	Proposition 3.32
Corollary 7.10 in the version of 2016-12-22	Corollary 7.53
Theorem 5.157 in the version of 15 February 2017	Theorem 6.160
Corollary 5.161 in the version of 25 May 2017	Corollary 6.164
Additional exercise 22 in the version of 25 May 2017	Exercise 6.56
Theorem 5.32 in the version of 25 May 2017	Theorem 6.32
Theorem 5.157 in the version of 25 May 2017	Theorem 6.160
Exercise 44 in the version of 28 October 2017	Exercise 6.22
Example 5.28 in the version of 7 November 2017	Example 6.7
Exercise 5.16 in the version of 18 November 2017	Exercise 6.17

item cited	its current location
Section 5.5 in the version of 14 March 2018	Section 6.5
Exercise 5.33 in the version of 21 March 2018	Exercise 6.33
Corollary 5.163 in the version of 21 March 2018	Corollary 6.164
Theorem 2.27 in the version of 26 April 2018	Theorem 3.30
Corollary 5.161 in the version of 26 April 2018	Corollary 6.162
Corollary 7.63 in the version of 26 April 2018	Corollary 7.70
Corollary 7.150 in the version of 26 April 2018	Corollary 7.182
Exercise 5.21 in the version of 14 September 2018	Exercise 6.21
Remark 2.27 in the version of 4 October 2018	Remark 2.27
§3.1 in the version of 4 October 2018	Section 3.1
Theorem 3.29 in the version of 4 October 2018	Theorem 3.29
Theorem 3.44 in the version of 4 October 2018	Theorem 3.44
Corollary 8.63 in the version of 4 October 2018	Corollary 7.69
Chapter 6 in the version of 18 October 2018	Chapter 6

References

[Abeles14] Francine F. Abeles, *Chiò's and Dodgson's determinantal identities*, Linear Algebra and its Applications, Volume 454, 1 August 2014, pp. 130–137.

[Aigner07] Martin Aigner, *A Course in Enumeration*, Graduate Texts in Mathematics #238, Springer 2007.

[AigZie14] Martin Aigner, Günter M. Ziegler, *Proofs from the Book*, 6th edition, Springer 2018.

[Aitken56] A. C. Aitken, *Determinants and Matrices*, 9th edition, Oliver and Boyd 1956.

[AleGhe14] Emily Allen, Irina Gheorghiciuc, *A weighted interpretation for the super Catalan numbers*, Journal of Integer Sequences, Vol. 17 (2014), Article 14.10.7.
A preprint is arXiv:1403.5246v2.

[AmaEsc05] Herbert Amann, Joachim Escher, *Analysis I*, translated from the German by Gary Brookfield, Birkhäuser 2005.

[AndDos10] Titu Andreescu, Gabriel Dospinescu, *Problems from the Book*, 2nd edition, XYZ Press 2010.

[AndDos12] Titu Andreescu, Gabriel Dospinescu, *Straight from the Book*, XYZ Press 2012.

[AndFen04] Titu Andreescu, Zuming Feng, *A Path to Combinatorics for Undergraduates: Counting Strategies*, Springer 2004.

[Artin10] Michael Artin, *Algebra*, 2nd edition, Pearson 2010.

[Axler15] Sheldon Axler, *Linear Algebra Done Right*, 3rd edition, Springer 2015.

[BenDre07] Arthur T. Benjamin and Gregory P. Dresden, *A Combinatorial Proof of Vandermonde's Determinant*, The American Mathematical Monthly, Vol. 114, No. 4 (Apr., 2007), pp. 338–341.
(Also available at `http://scholarship.claremont.edu/hmc_fac_pub/524/` .)

[BenQui03] Arthur T. Benjamin and Jennifer J. Quinn, *Proofs that Really Count: The Art of Combinatorial Proof*, The Mathematical Association of America, 2003.

[BenQui04] Arthur T. Benjamin and Jennifer J. Quinn, *Proofs that Really Count: The Magic of Fibonacci Numbers and More*, Mathematical Adventures for Students and Amateurs, (David F. Hayes and Tatiana Shubin, editors), Spectrum Series of MAA, pp. 83–98, 2004.

[BenQui08] Arthur T. Benjamin and Jennifer J. Quinn, *An Alternate Approach to Alternating Sums: A Method to DIE for*, The College Mathematics Journal, Volume 39, Number 3, May 2008, pp. 191-202(12).

[BerBru08] Adam Berliner and Richard A. Brualdi, *A combinatorial proof of the Dodgson/Muir determinantal identity*, International Journal of Information and Systems Sciences, Volume 4 (2008), Number 1, pp. 1–7.

[Bergma15] George M. Bergman, *A Companion to Lang's Algebra*, website (2015).
`https://math.berkeley.edu/~gbergman/.C.to.L/`

[BirMac99] Saunders Mac Lane, Garrett Birkhoff, *Algebra*, 3rd edition, AMS Chelsea Publishing 1999.

[BjoBre05] Anders Björner, Francesco Brenti, *Combinatorics of Coxeter Groups*, Graduate Texts in Mathematics #231, Springer 2005.

[Bona12] Miklós Bóna, *Combinatorics of Permutations*, 2nd edition, CRC Press 2012.

[Bourba74] Nicolas Bourbaki, *Algebra I, Chapters 1-3*, Hermann 1974.
`https://archive.org/details/ElementsOfMathematics-AlgebraPart1/`

[Bresso99] David M. Bressoud, *Proofs and Confirmations: The Story of the Alternating Sign Matrix Conjecture*, Cambridge University Press 1999.
See `https://www.macalester.edu/~bressoud/books/PnC/PnCcorrect.html` for errata.

[BruRys91] Richard A. Brualdi, Herbert J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press 1991.

[Bump02] Daniel Bump, *Mathematics of the Rubik's Cube*, lecture notes (in 2 versions).
`http://sporadic.stanford.edu/bump/match/rubik.html`

[CamFon07] Peter J. Cameron, Dima G. Fon-Der-Flaass, *Fibonacci notes*, 1 August 2007.
`http://www.maths.qmul.ac.uk/~pjc/comb/fibo.pdf`

[CaSoSp12] Sergio Caracciolo, Alan D. Sokal, Andrea Sportiello, *Algebraic/combinatorial proofs of Cayley-type identities for derivatives of determinants and pfaffians*, Advances in Applied Mathematics 50, pp. 474–594 (2013). Also appears as arXiv preprint `arXiv:1105.6270v2`.

[Comtet74] Louis Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions*, D. Reidel Publishing Company, 1974.

[Conrad1] Keith Conrad, *Sign of permutations*,
`https://kconrad.math.uconn.edu/blurbs/grouptheory/sign.pdf` .

[Conrad2] Keith Conrad, *Bilinear forms*,
`https://kconrad.math.uconn.edu/blurbs/linmultialg/bilinearform.pdf`

[Conrad3] Keith Conrad, *Conjugation in a group*.
`https://kconrad.math.uconn.edu/blurbs/grouptheory/conjclass.pdf`

[DanRot78] Ottavio D'Antona, Gian-Carlo Rota, *Two Rings Connected with the Inclusion-Exclusion Principle*, Journal of Combinatorial Theory, Series A, Volume 24, Issue 3, May 1978, pp. 395–402.

[daSilv12] Patrick Da Silva, *Polynomial in* $\mathbb{Q}[x]$ *sending integers to integers?, math.stackexchange answer #108318*.

[Day16] Martin V. Day, *An Introduction to Proofs and the Mathematical Vernacular*, 7 December 2016.
`https://web.archive.org/web/20180521211821/https://www.math.vt.edu/people/day/ProofsBook/IPaMV.pdf` .

[DiaGra77] Persi Diaconis, R. L. Graham, *Spearman's Footrule as a Measure of Disarray*, Journal of the Royal Statistical Society, Series B **39**, No. 2 (1977), pp. 262–268.
`https://statweb.stanford.edu/~cgates/PERSI/papers/77_04_spearmans.pdf`

[dilemi17] Dilemian et al, *math.stackexchange post #2455428 ("An identity involving binomial coefficients and rational functions")*, `https://math.stackexchange.com/q/2455428` .

[DouNie19] Darrin Doud and Pace P. Nielsen, *A Transition to Advanced Mathematics*, 10 June 2019.
`https://math.byu.edu/~pace/Transition-ereader.pdf`

[EdeStr04] Alan Edelman and Gilbert Strang, *Pascal Matrices*, American Mathematical Monthly, Vol. 111, No. 3 (March 2004), pp. 189–197.

[Eisenk99] Theresia Eisenkölbl, *Rhombus Tilings of a Hexagon with Three Fixed Border Tiles*, arXiv:math/9712261v2. Published in: J. Combin. Theory Ser. A 88 (1999), pp. 368–378.

[Elman18] Richard Elman, *Lectures on Abstract Algebra*, 20 September 2019.
`https://www.math.ucla.edu/~rse/algebra_book.pdf`

[FadSom72] D. Faddeev, I. Sominsky, *Problems in Higher Algebra*, Mir 1972.

[Fische01] Hanspeter Fischer, *Linear recurrence relations with constant coefficients*, 9 April 2001.
`http://matematicas.uam.es/~mavi.melian/CURSO_15_16/web_Discreta/recurrence.pdf`

[FomZel01] Sergey Fomin, Andrei Zelevinsky, *Cluster Algebras I: Foundations*, Journal of the American Mathematical Society, Volume 15, Number 2, pp. 497–529.
`https://doi.org/10.1090/S0894-0347-01-00385-X`

[FomZel02] Sergey Fomin, Andrei Zelevinsky, *The Laurent Phenomenon*, Advances in Applied Mathematics **28** (2002), pp. 119–144.

[FoWiZe16] Sergey Fomin, Lauren Williams, Andrei Zelevinsky, *Introduction to Cluster Algebras. Chapters 1-3*, arXiv:1608.05735v2.

[Fulton97] William Fulton, *Young Tableaux: With Applications to Representation Theory and Geometry*, Cambridge University Press 1997.

[Gale98] David Gale, *Tracking the automatic ant and other mathematical explorations*, Springer 1998.

[GalKar71] David Gale, Richard M. Karp, *A phenomenon in the theory of sorting*, Journal of Computer and System Sciences, Volume 6, Issue 2, April 1972, pp. 103–115.

[GalQua18] Jean Gallier and Jocelyn Quaintance, *Algebra, Topology, Differential Calculus, and Optimization Theory For Computer Science and Engineering*, 2 August 2019.
https://www.cis.upenn.edu/~jean/gbooks/geomath.html

[Galvin17] David Galvin, *Basic discrete mathematics*, 13 December 2017.
http://www-users.math.umn.edu/~dgrinber/comb/60610lectures2017-Galvin.pdf
(The URL might change, and the text may get updated. In order to reliably obtain the version of 13 December 2017, you can use the archive.org Wayback Machine: https://web.archive.org/web/20180205122609/http://www-users.math.umn.edu/~dgrinber/comb/60610lectures2017-Galvin.pdf .)

[Gessel79] Ira Gessel, *Tournaments and Vandermonde's Determinant*, Journal of Graph Theory, Vol. 3 (1979), pp. 305–307.

[Gessel92] Ira M. Gessel, *Super Ballot Numbers*, Journal of Symbolic Computation, Volume 14, Issues 2–3, August-September 1992, pp. 179–194.
https://doi.org/10.1016/0747-7171(92)90034-2

[GohKol96] Israel Gohberg, Israel Koltracht, *Triangular factors of Cauchy and Vandermonde matrices*, Integr. Equat. Oper. Th. Vol. 26 (1996), pp. 46–59.

[Goodma15] Frederick M. Goodman, *Algebra: Abstract and Concrete*, edition 2.6, 1 May 2015.
http://homepage.math.uiowa.edu/~goodman/algebrabook.dir/book.2.6.pdf .

[Gould10] H. W. Gould, *Combinatorial Identities: Table I: Intermediate Techniques for Summing Finite Series*, Edited and Compiled by Jocelyn Quaintance, May 3, 2010.

[Grinbe09] Darij Grinberg, *Solution to Problem 19.9 from "Problems from the Book"*.
http://www.cip.ifi.lmu.de/~grinberg/solutions.html

[Grinbe10] Darij Grinberg, *A hyperfactorial divisibility*, version of 27 July 2015.
http://www.cip.ifi.lmu.de/~grinberg/

[Grinbe11] Darij Grinberg, *Zeckendorf family identities generalized*, arXiv preprint arXiv:1103.4507v2.

[Grinbe16a] Darij Grinberg, *4th QEDMO, Problem 13 with solution*.

[Grinbe16b] Darij Grinberg, *Notes on linear algebra*, version of 26 September 2019. `https://github.com/darijgr/lina`

[GriRei18] Darij Grinberg, Victor Reiner, *Hopf algebras in Combinatorics*, version of 11 May 2018, `arXiv:1409.8356v5`.
See also `http://www.cip.ifi.lmu.de/~grinberg/algebra/HopfComb-sols.pdf` for a version that gets updated.

[GrKnPa94] Ronald L. Graham, Donald E. Knuth, Oren Patashnik, *Concrete Mathematics, Second Edition*, Addison-Wesley 1994.
See `https://www-cs-faculty.stanford.edu/~knuth/gkp.html` for errata.

[Hammac15] Richard Hammack, *Book of Proof*, 3rd edition 2018. `https://www.people.vcu.edu/~rhammack/BookOfProof/`

[Heffer17] Jim Hefferon, *Linear Algebra*, 3rd edition, 1 January 2017, `http://joshua.smcvt.edu/linearalgebra/` .

[Heinig11] Peter Christian Heinig, *Chio Condensation and Random Sign Matrices*, arXiv:1103.2717v3.

[Herstei75] I. N. Herstein, *Topics in Algebra*, 2nd edition, John Wiley & Sons, 1975.

[HofKun71] Kenneth Hoffman, Ray Kunze, *Linear algebra*, 2nd edition, Prentice-Hall 1971.
See `https://math.stackexchange.com/questions/437253/` for errata.

[Hunger03] Thomas W. Hungerford, *Algebra*, 12th printing, Springer 2003.

[Hunger14] Thomas W. Hungerford, *Abstract Algebra: An Introduction*, 3rd edition, Brooks/Cole 2014.

[Jacobs10] Nathan Jacobson, *Finite-Dimensional Division Algebras Over Fields*, corrected 2nd printing, Springer 2010.

[Joyce17] David Joyce, *Introduction to Modern Algebra*, 5 December 2017. `https://mathcs.clarku.edu/~djoyce/ma225/algebra.pdf`

[Joyner08] W. D. Joyner, *Mathematics of the Rubik's cube*, 19 August 2008. `https://web.archive.org/web/20160304122348/http://www.permutationpuzzles.org/rubik/webnotes/` (link to the PDF at the bottom).

[KarZha16] Karthik Karnik, Anya Zhang, *Combinatorial proof of Chio Pivotal Condensation*, 25 May 2016.

[KenWil14] Richard W. Kenyon, David B. Wilson, *The space of circular planar electrical networks*, arXiv:1411.7425v3.

[Kerber99] Adalbert Kerber, *Applied Finite Group Actions*, 2nd edition, Springer 1999.

[KleLak72] S. L. Kleiman and Dan Laksov, *Schubert Calculus*, The American Mathematical Monthly, Vol. 79, No. 10, (Dec., 1972), pp. 1061–1082.

[Knapp16] Anthony W. Knapp, *Basic Algebra*, digital second edition 2016. `http://www.math.stonybrook.edu/~aknapp/download.html`

[Knuth88] Donald E. Knuth, *Fibonacci Multiplication*, Appl. Math. Lett., Vol. 1, No. 1, pp. 57–60, 1988.

[Knutso12] Allen Knutson, *Schubert polynomials and Symmetric Functions*, lecture notes, July 28, 2012.

[Kratt99] Christian Krattenthaler, *Advanced Determinant Calculus*, Séminaire Lotharingien Combin. 42 (1999) (The Andrews Festschrift), paper B42q, 67 pp., arXiv:math/9902004v3.

[KurLis78] L. Kurlyandchik, A. Lisitskiy, *Summy i proizvedeniya ("Sums and products")*, Kvant 1978, issue 10, pp. 31–37. `http://kvant.mccme.ru/1978/10/summy_i_proizvedeniya.htm`

[LLPT95] D. Laksov, A. Lascoux, P. Pragacz, and A. Thorup, *The LLPT Notes*, edited by A. Thorup, 1995, `http://www.math.ku.dk/~thorup/notes/sympol.pdf` .

[Lalond96] Pierre Lalonde, *A non-commutative version of Jacobi's equality on the cofactors of a matrix*, Discrete Mathematics 158 (1996), pp. 161–172. `https://doi.org/10.1016/0012-365X(95)00040-4` .

[Lampe13] Philipp Lampe, *Cluster algebras*, 4 December 2013, `http://www.math.uni-bielefeld.de/~lampe/teaching/cluster/cluster.pdf` .

[Lang02] Serge Lang, *Algebra*, Revised Third Edition, Graduate Texts in Mathematics #211, Springer 2002.

[Laue15] Hartmut Laue, *Determinants*, version 17 May 2015, `http://www.math.uni-kiel.de/algebra/laue/homepagetexte/det.pdf` .

[LeeSch11] Kyungyong Lee, Ralf Schiffler, *A Combinatorial Formula for Rank* 2 *Cluster Variables*, Journal of Algebraic Combinatorics **37** (2013), Issue 1, pp. 67–85.
A preprint is available at `http://arxiv.org/abs/1106.0952v3` .

[LeeSch13] Kyungyong Lee, Ralf Schiffler, *Positivity for cluster algebras*, Annals of Mathematics (second series) **182**, No. 1 (July, 2015), pp. 73–125.
A preprint is available at `http://arxiv.org/abs/1306.2415v3` .

[Leeuwe06] Marc A. A. van Leeuwen, *Schur functions and alternating sums*, Electronic Journal of Combinatorics Vol 11(2) A5 (2006), also available as arXiv:math.CO/0602357.

[LeLeMe16] Eric Lehman, F. Thomson Leighton, Albert R. Meyer, *Mathematics for Computer Science*, revised Tuesday 6th June 2018,
`https://courses.csail.mit.edu/6.042/spring18/mcs.pdf` .

[Loehr11] Nicholas A. Loehr, *Bijective Combinatorics*, Chapman & Hall/CRC 2011.

[Manive01] Laurent Manivel, *Symmetric Functions, Schubert Polynomials and Degeneracy Loci*, SMF/AMS Texts and Monographs #6, American Mathematical Society 2001.

[Marsh13] Robert J. Marsh, *Lecture Notes on Cluster Algebras*, European Mathematical Society 2013.
See `http://www1.maths.leeds.ac.uk/~marsh/errata.pdf` for a list of errata.

[Mate14] Attila Maté, *Determinants*, version 1 October 2017.
`http://www.sci.brooklyn.cuny.edu/~mate/misc/determinants.pdf`

[Muir30] Thomas Muir, *The theory of determinants in the historical order of development*, 5 volumes (1906–1930), later reprinted by Dover.
`http://www-igm.univ-mlv.fr/~al/`

[MuiMet60] Thomas Muir, *A Treatise on the Theory of Determinants*, revised and enlarged by William H. Metzler, Dover 1960.

[Mulhol16] Jamie Mulholland, *Permutation Puzzles: A Mathematical Perspective*,
`http://www.sfu.ca/~jtmulhol/math302/`

[MusPro07] Gregg Musiker, James Propp, *Combinatorial Interpretations for Rank-Two Cluster Algebras of Affine Type*, The Electronic Journal of Combinatorics **14** (2007), #R15.
`http://www.combinatorics.org/ojs/index.php/eljc/article/view/v14i1r15`

[Newste19] Clive Newstead, *An Infinite Descent into Pure Mathematics*, version 0.4 (1 January 2020).
`https://infinitedescent.xyz`

[NouYam02] Masatoshi Noumi, Yasuhiko Yamada, *Tropical Robinson-Schensted-Knuth correspondence and birational Weyl group actions*, arXiv:math-ph/0203030v2.

[OlvSha18] Peter J. Olver, Chehrzad Shakiban, *Applied Linear Algebra*, 2nd edition, Springer 2018.
`https://doi.org/10.1007/978-3-319-91041-3`
See `http://www.math.umn.edu/~olver/ala.html` for errata.

[OruPhi00] Halil Oruç, George M. Phillips, *Explicit factorization of the Vandermonde matrix*, Linear Algebra and its Applications 315 (2000), pp. 113–123.

[PetTen14] T. Kyle Petersen, Bridget Eileen Tenner, *The depth of a permutation*, Journal of Combinatorics **6** (2015), Number 1–2, pp. 145–178, arXiv:1202.4765v3.

[Pinkha15] Henry C. Pinkham, *Linear Algebra*, draft of a textbook, version 10 July 2015.
`https://www.math.columbia.edu/~pinkham/HCP_LinearAlgebra.pdf`

[Prasol94] Viktor V. Prasolov, *Problems and Theorems in Linear Algebra*, Translations of Mathematical Monographs, vol. #134, AMS 1994.

[Richma88] Fred Richman, *Nontrivial uses of trivial rings*, Proceedings of the American Mathematical Society, Volume 103, Number 4, August 1988, pp. 1012–1014.

[Riorda68] John Riordan, *Combinatorial Identities*, Robert E. Krieger Publishing Company, Huntington, NY 1979.

[Rocket81] Andrew M. Rockett, *Sums of the Inverses of Binomial Coefficients*, The Fibonacci Quarterly 19 (1981), issue 5, pp. 433–437.
`https://www.fq.math.ca/Scanned/19-5/rockett.pdf`

[Rote01] Günter Rote, *Division-Free Algorithms for the Determinant and the Pfaffian: Algebraic and Combinatorial Approaches*, Computational Discrete Mathematics, Lecture Notes in Computer Science, Volume 2122, 2001, pp. 119–135.
`https://web.archive.org/web/20161005071023/http://www.inf.fu-berlin.de/groups/ag-ti/people/rote/Papers/pdf/Division-free+algorithms.pdf`

[Rotman15] Joseph J. Rotman, *Advanced Modern Algebra: Part 1*, Graduate Studies in Mathematics #165, 3rd edition, AMS 2015.

[SacUlf11] Joshua Sack, Henning Úlfarsson, *Refined inversion statistics on permutations*, The Electronic Journal of Combinatorics **19** (2012), #P29. See also arXiv:1106.1995v2 for a preprint version.

[Schmit04] William R. Schmitt, *Incidence Hopf algebras*, Journal of Pure and Applied Algebra **96** (1994), pp. 299–330, `https://doi.org/10.1016/0022-4049(94)90105-8` .

[Silves00] John R. Silvester, *Determinants of Block Matrices*, The Mathematical Gazette, Vol. 84, No. 501 (Nov., 2000), pp. 460–467.

[Spivey12a] Mike Spivey, *math.stackexchange post #180440 ("Alternating sum of squares of binomial coefficients")*, `https://math.stackexchange.com/a/180440/` .

[Spivey12b] Mike Spivey, *Combinatorial Interpretation of the Alternating Convolution of the Central Binomial Coefficients*, `https://mikespivey.wordpress.com/2012/03/16/altconvcentralbinom/` .

[Stanle11] Richard Stanley, *Enumerative Combinatorics, volume 1*, Second edition, Cambridge University Press 2012.
See `http://math.mit.edu/~rstan/ec/` for errata and for a preprint version (from 2011).

[Stanle01] Richard Stanley, *Enumerative Combinatorics, volume 2*, First edition, Cambridge University Press 2001.
See `http://math.mit.edu/~rstan/ec/` for errata.

[Stembr02] John R. Stembridge, *A Concise Proof of the Littlewood-Richardson Rule*, The Electronic Journal of Combinatorics, Volume 9 (2002), Note #N5.

[Strick13] Neil Strickland, *MAS201 Linear Mathematics for Applications*, lecture notes, 28 September 2013.
`http://neil-strickland.staff.shef.ac.uk/courses/MAS201/`
See `http://www.cip.ifi.lmu.de/~grinberg/t/19fla/MAS201.pdf` for a corrected and expanded version of these notes (as of 1 December 2019).

[Sved84] Marta Sved, *Counting and Recounting: The Aftermath*, The Mathematical Intelligencer **6** (1984), no. 2, pp. 44–45.

[Swanso18] Irena Swanson, *Introduction to Analysis*, with construction of the number systems, 18 November 2019.
`http://people.reed.edu/~iswanson/analysisconstructR.pdf`

[Tenner04] Bridget Eileen Tenner, *A Non-Messing-Up Phenomenon for Posets*, Annals of Combinatorics 11 (2007), pp. 101–114.
A preprint is available as arXiv:math/0404396.

[Turnbu29] H. W. Turnbull, *The Theory of Determinants, Matrices, and Invariants*, Blackie & Son Ltd, 1929.
`https://archive.org/details/in.ernet.dli.2015.4951`

[Vellem06] Daniel J. Velleman, *How to Prove It*, 2nd edition, Cambridge University Press 2006.

[Vorobi02] Nicolai N. Vorobiev, *Fibonacci Numbers*, Translated from the Russian by Mircea Martin, Springer 2002 (translation of the 6th Russian edition).

[Walker87] Elbert A. Walker, *Introduction to Abstract Algebra*, Random House/Birkhauser, New York, 1987.

[Willia03] Geordie Williamson, *Mind your P and Q-symbols: Why the Kazhdan-Lusztig basis of the Hecke algebra of type A is cellular*, Honours thesis at the University of Sydney, October 2003.

[Willia15] Stanley Gill Williamson, *The common-submatrix Laplace expansion*, arXiv:1505.05486v1.

[Willia18] Stanley Gill Williamson, *Matrix Canonical Forms: notational skills and proof techniques*, 28 April 2018.

[Zeilbe85] Doron Zeilberger, *A combinatorial approach to matrix algebra*, Discrete Mathematics 56 (1985), pp. 61–72.

[Zeilbe93] Doron Zeilberger, *On an identity of Daubechies*, The American Mathematical Monthly, Vol. 100, No. 5 (May, 1993), p. 487.
`http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/wavelet.html`

[Zeilbe98] Doron Zeilberger, *Dodgson's Determinant-Evaluation Rule proved by Two-Timing Men and Women*, The Electronic Journal of Combinatorics, vol. 4, issue 2 (1997) (The Wilf Festschrift volume), R22.
`http://www.combinatorics.org/ojs/index.php/eljc/article/view/v4i2r22`
Also available as arXiv:math/9808079v1.
`http://arxiv.org/abs/math/9808079v1`