Linear recurrences with constant coefficients: the multivariate case

MIREILLE BOUSQUET-MÉLOU*
LaBRI, Université Bordeaux 1
351 cours de la Libération
33405 Talence Cedex, France
bousquet@labri.u-bordeaux.fr

MARKO PETKOVŠEK†
Dept. Mathematics and Mechanics
University of Ljubljana
Jadranska 19, 1111 Ljubljana, Slovenia
Marko.Petkovsek@fmf.uni-lj.si

Abstract

While in the univariate case solutions of linear recurrences with constant coefficients have rational generating functions, we show that the multivariate case is much richer: even though initial conditions have rational generating functions, the corresponding solutions can have generating functions which are algebraic but not rational, D-finite but not algebraic, and even non D-finite.

1 Introduction

The aim of this paper is to study the nature of multivariate generating functions

$$F(x_1, \ldots, x_d) = \sum_{n_1,\ldots,n_d \geq 0} a_{n_1,\ldots,n_d} x_1^{n_1} \cdots x_d^{n_d} = \sum_{\boldsymbol{n} \geq \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}}$$

whose coefficients satisfy a linear recurrence relation with constant coefficients

$$a_{\boldsymbol{n}} = c_{\boldsymbol{h}_1} a_{\boldsymbol{n}+\boldsymbol{h}_1} + c_{\boldsymbol{h}_2} a_{\boldsymbol{n}+\boldsymbol{h}_2} + \cdots + c_{\boldsymbol{h}_k} a_{\boldsymbol{n}+\boldsymbol{h}_k} \quad \text{for } \boldsymbol{n} \geq \boldsymbol{s}\,,$$

with adequate initial conditions. The univariate case ($d = 1$) is well-known to give rise to rational generating functions. We shall show that the multivariate case is much richer.

We start by an existence and uniqueness theorem that actually applies to recurrences of a more general form

$$a_{\boldsymbol{n}} = \Phi(a_{\boldsymbol{n}+\boldsymbol{h}_1}, a_{\boldsymbol{n}+\boldsymbol{h}_2}, \ldots, a_{\boldsymbol{n}+\boldsymbol{h}_k}) \quad \text{for } \boldsymbol{n} \geq \boldsymbol{s}\,,$$

where the values of $a_{\boldsymbol{n}}$ for $\boldsymbol{n} \not\geq \boldsymbol{s}$ are given explicitly. In particular, we show that the lattice points in the first orthant (*i.e.*, $\boldsymbol{n} \geq \boldsymbol{0}$) can be enumerated in such a way that for all $\boldsymbol{n}$

*Partially supported by the Conseil Régional d'Aquitaine.
†Partially supported by MZT RS under grant J2-8549.

the points $\boldsymbol{n}+\boldsymbol{h}_i$ precede $\boldsymbol{n}$ in this enumeration, if and only if the convex hull of the set $H = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_k\}$ does not intersect the first orthant (Section 2). This condition on H, which ensures existence and uniqueness of the solution, is assumed to be satisfied in the sequel.

For linear recurrences with constant coefficients, we show that when initial conditions grow at most exponentially, the same is true of the solution, which is consequently analytic in a neighbourhood of the origin (Section 3).

Next, we study the algebraic nature of the generating function of the solution of such recurrences (Section 4). We define the *apex* of H as the componentwise maximum of the points in $H \cup \{\mathbf{0}\}$. When the initial conditions have rational generating functions and the apex of H is $\mathbf{0}$, the generating function of the solution is rational and given by an explicit formula. When the initial conditions have algebraic generating functions and the apex of H has at most one positive coordinate, the generating function of the solution is algebraic and is given explicitly in terms of the solutions of an algebraic equation. We give various applications to the enumeration of lattice paths that generalize Dyck paths.

When the apex has more than one positive coordinate, and the initial conditions have rational generating functions, the generating function of the solution need not be algebraic, which we demonstrate on the problem of Young tableaux of bounded height. However, the solution here remains D-finite. We conclude the paper with a simple example showing that non-D-finite solutions (and even, hypertranscendental solutions) might also occur.

The main tool that we use is the so-called *kernel method* which permits to solve certain systems of linear functional equations that seem to contain too many “unknowns”. It works by restricting the equations to algebraic varieties on which some of the unknown terms vanish, thus providing the “missing” equations (see Section 4 for more detail).

Notations. We use $\mathbb{N}$ to denote the set of nonnegative integers. We write $\boldsymbol{u} = (u_1, u_2, \ldots, u_d)$ for d-tuples of numbers or indeterminates, $\mathbf{0} = (0, 0, \ldots, 0)$, $\mathbf{1} = (1, 1, \ldots, 1)$, $\boldsymbol{u} \geq \boldsymbol{v}$ when $u_i \geq v_i$ for $1 \leq i \leq d$, and $\boldsymbol{u} > \boldsymbol{v}$ when $u_i > v_i$ for $1 \leq i \leq d$. The monomial $x_1^{u_1} \cdots x_d^{u_d}$ is denoted $\boldsymbol{x}^{\boldsymbol{u}}$. For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{Z}^d$, the scalar product $u_1 v_1 + \cdots + u_d v_d$ is denoted $\boldsymbol{u} \cdot \boldsymbol{v}$. The convex hull of a set $H \subseteq \mathbb{R}^d$ is denoted $\operatorname{conv} H$.

2 An existence and uniqueness theorem

Let A be a nonempty set. We consider d-dimensional recurrence equations of the form

$$a_{\boldsymbol{n}} = \Phi(a_{\boldsymbol{n}+\boldsymbol{h}_1}, a_{\boldsymbol{n}+\boldsymbol{h}_2}, \ldots, a_{\boldsymbol{n}+\boldsymbol{h}_k}) \qquad \text{for } \boldsymbol{n} \geq \boldsymbol{s}, \tag{1}$$

where $a : \mathbb{N}^d \to A$ is the unknown d-dimensional sequence of elements of A, $\Phi : A^k \to A$ is a given function, $H = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_k\} \subseteq \mathbb{Z}^d$ is the set of *shifts*, and $\boldsymbol{s} \in \mathbb{N}^d$ is the *starting point* satisfying $\boldsymbol{s} + H \subseteq \mathbb{N}^d$. A given function φ specifies the initial conditions:

$$a_{\boldsymbol{n}} = \varphi(\boldsymbol{n}) \qquad \text{for } \boldsymbol{n} \geq \mathbf{0},\ \boldsymbol{n} \not\geq \boldsymbol{s}\,. \tag{2}$$

We think of the $\boldsymbol{h}_i$ as having mostly (but not necessarily only) negative coordinates, and of the point $\boldsymbol{n}$ as depending on the points $\boldsymbol{n}+\boldsymbol{h}_1, \boldsymbol{n}+\boldsymbol{h}_2, \ldots, \boldsymbol{n}+\boldsymbol{h}_k$ as far as the value of $a_{\boldsymbol{n}}$ is concerned.

The objective of this section is to characterize the sets H for which there is an ordering of $\mathbb{N}^d$ of order type ω such that the points $\boldsymbol{n}+\boldsymbol{h}_1, \boldsymbol{n}+\boldsymbol{h}_2, \ldots, \boldsymbol{n}+\boldsymbol{h}_k$ precede $\boldsymbol{n}$ in this ordering. Then there exists a unique solution of (1) – (2), and for any $\boldsymbol{n} \in \mathbb{N}^d$ it is possible to compute the value of $a_{\boldsymbol{n}}$ directly from (1) – (2) in a finite number of steps.

We first define a dependency relation on the points of $\mathbb{N}^d$, induced by the set H.

Definition 1 *For $H \subseteq \mathbb{Z}^d$ and $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{N}^d$, let*

$$\boldsymbol{p} \prec_H \boldsymbol{q} \quad \textit{if and only if} \quad \boldsymbol{p} \in \boldsymbol{q} + H \subseteq \mathbb{N}^d .$$

The transitive closure $\prec_H^+$ of $\prec_H$ in $\mathbb{N}^d$ is the dependency relation *corresponding to H.*

Lemma 2 *Let $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{N}^d$. Then*

$$\boldsymbol{p} \prec_H^+ \boldsymbol{q},\ \boldsymbol{u} \prec_H^+ \boldsymbol{v} \quad \textit{implies} \quad \boldsymbol{p}+\boldsymbol{u} \prec_H^+ \boldsymbol{q}+\boldsymbol{v} .$$

Proof: Clearly $\prec_H$ is translation-invariant: if $\boldsymbol{p} \prec_H \boldsymbol{q}$, then $\boldsymbol{p}+\boldsymbol{r} \prec_H \boldsymbol{q}+\boldsymbol{r}$. Therefore $\prec_H^+$ is translation-invariant as well, so $\boldsymbol{p} \prec_H^+ \boldsymbol{q}$ implies $\boldsymbol{p}+\boldsymbol{u} \prec_H^+ \boldsymbol{q}+\boldsymbol{u}$ and $\boldsymbol{u} \prec_H^+ \boldsymbol{v}$ implies $\boldsymbol{q}+\boldsymbol{u} \prec_H^+ \boldsymbol{q}+\boldsymbol{v}$. By transitivity, $\boldsymbol{p}+\boldsymbol{u} \prec_H^+ \boldsymbol{q}+\boldsymbol{v}$. ∎

Recall that a binary relation $R \subseteq A \times A$ is *well-founded* if it is transitive and has no infinite descending chains in A (there exists no sequence $(a_n)_{n\ge 1}$ of elements of A such that $a_{n+1}\ R\ a_n$ for all n). Note that a well-founded relation is irreflexive and asymmetric.

Theorem 3 *Let $H \subseteq \mathbb{Z}^d$ be a finite set, and $\prec_H^+$ the corresponding dependency relation. Then the following are equivalent:*

(i) *$\prec_H^+$ is well-founded in $\mathbb{N}^d$,*

(ii) $\{\boldsymbol{x} \in \mathbb{R}^d;\ \boldsymbol{x} \ge \boldsymbol{0}\} \ \cap\ \operatorname{conv} H \ =\ \emptyset$,

(iii) *there exists $\boldsymbol{v} \in \mathbb{N}^d$, $\boldsymbol{v} > \boldsymbol{0}$, such that $\boldsymbol{v} \cdot \boldsymbol{h} < 0$ for all $\boldsymbol{h} \in H$,*

(iv) *$\prec_H^+$ can be extended to an ordering of $\mathbb{N}^d$ of order type ω.*

First we establish a lemma about *rational points* in convex subsets of $\mathbb{R}^d$ (a point is *rational* if it belongs to $\mathbb{Q}^d$).

Lemma 4 *Let $X \subseteq \mathbb{R}^d$ be a finite set of rational points, and $\boldsymbol{p} \in \operatorname{conv} X$. Then $\boldsymbol{p}$ is rational if and only if it can be expressed as a rational convex combination of points from X.*

Proof: If $\boldsymbol{p}$ is a rational convex combination of rational points then clearly $\boldsymbol{p}$ is rational.

Conversely, let $\boldsymbol{p} \in \operatorname{conv} X$ be rational. Let $S \subseteq X$ be a minimal subset with the property that $\boldsymbol{p} \in \operatorname{conv} S$. Let e be the affine dimension of S. By Carathéodory's theorem (*cf.* [21, Thm. 17.1]) and by minimality of S, we have $|S| \le e+1$. But $e+1$ equals the maximum number of affinely independent points in S, by definition of e. So $|S| = e+1$ and S is affinely independent. Therefore $\operatorname{conv} S$ is an e-simplex containing $\boldsymbol{p}$.

Let $S = \{\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_e\}$ and $S_i = S \cup \{\boldsymbol{p}\} \setminus \{\boldsymbol{q}_i\}$, for $i = 0, 1, \ldots, e$. Then $\boldsymbol{p} = \sum_{i=0}^{e} \lambda_i\, \boldsymbol{q}_i$ where

$$\lambda_i = \frac{e\text{-vol}(\operatorname{conv} S_i)}{e\text{-vol}(\operatorname{conv} S)}\,, \qquad \text{for } i = 0, 1, \ldots, e$$

(see, *e.g.*, [11]). Since the volume of a simplex is a fraction of a determinant whose nonunit entries are the coordinates of its vertices, any simplex with rational vertices has rational volume. It follows that all the λ_i are rational. ■

Proof of Theorem 3: If H is empty then $\prec_H^+$ is the empty relation and all four assertions are trivially true. Now assume that H is nonempty.

(i) ⇒ (ii) Let $K = \{\boldsymbol{x} \in \mathbb{R}^d;\ \boldsymbol{x} \geq \boldsymbol{0}\} \cap \operatorname{conv} H$. Assume that K is nonempty. Then K is a convex polytope with rational vertices. Let $\boldsymbol{q} \in K \cap \mathbb{Q}^d$. By Lemma 4, $\boldsymbol{q}$ is a rational convex combination of points from H. Let N be the least common denominator of the coefficients in this combination, and $\boldsymbol{z} = N\boldsymbol{q}$. Then $\boldsymbol{z} \geq \boldsymbol{0}$ and $\boldsymbol{z} = \sum_{\boldsymbol{h} \in H} \mu_{\boldsymbol{h}}\, \boldsymbol{h}$ where $\mu_{\boldsymbol{h}} \in \mathbb{N}$ are not all zero. Let $\boldsymbol{s} \in \mathbb{N}^d$ be such that $\boldsymbol{s} + H \subseteq \mathbb{N}^d$. Then $\boldsymbol{s} + \boldsymbol{h} \prec_H \boldsymbol{s}$ for all $\boldsymbol{h} \in H$. Let $M = \sum_{\boldsymbol{h} \in H} \mu_{\boldsymbol{h}}$. By repeated application of Lemma 2, $\mu_{\boldsymbol{h}}\, \boldsymbol{s} + \mu_{\boldsymbol{h}}\, \boldsymbol{h} \prec_H^+ \mu_{\boldsymbol{h}}\, \boldsymbol{s}$ for all $\boldsymbol{h} \in H$, and $M\boldsymbol{s} + \boldsymbol{z} \prec_H^+ M\boldsymbol{s}$. Let $\boldsymbol{z}_k = M\boldsymbol{s} + k\,\boldsymbol{z}$. Note that $\boldsymbol{z}_k \in \mathbb{N}^d$ for all $k \in \mathbb{N}$. Using Lemma 2 it follows by induction on k that $\boldsymbol{z}_{k+1} \prec_H^+ \boldsymbol{z}_k$ for all $k \in \mathbb{N}$. Thus the $\boldsymbol{z}_k$ form an infinite descending chain in $\mathbb{N}^d$, contradicting the well-foundedness of $\prec_H^+$. Hence $K = \emptyset$.

(ii) ⇒ (iii) As $\{\boldsymbol{x} \in \mathbb{R}^d;\ \boldsymbol{x} \geq \boldsymbol{0}\}$ and $\operatorname{conv} H$ are disjoint convex polyhedra they can be separated by a hyperplane which meets neither of them. Therefore there are $\boldsymbol{u} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ and $b \in \mathbb{R}$ such that $\boldsymbol{u} \cdot \boldsymbol{x} < b$ for all $\boldsymbol{x} \in \operatorname{conv} H$ and $\boldsymbol{u} \cdot \boldsymbol{x} > b$ for all $\boldsymbol{x} \geq \boldsymbol{0}$. In particular, $\boldsymbol{u} \cdot \boldsymbol{h} < b < 0$ for all $\boldsymbol{h} \in H$. If $u_i < 0$ for some i then $\boldsymbol{x} = (0, \ldots, 0, b/u_i, 0, \ldots, 0) \geq \boldsymbol{0}$ but $\boldsymbol{u} \cdot \boldsymbol{x} = b \not> b$. It follows that $\boldsymbol{u} \geq \boldsymbol{0}$.

As H is finite, and $\{\boldsymbol{w} \in \mathbb{Q}^d :\ \boldsymbol{w} > \boldsymbol{0}\}$ is dense in $\{\boldsymbol{u} \in \mathbb{R}^d :\ \boldsymbol{u} \geq \boldsymbol{0}\}$, we can change $\boldsymbol{u}$ slightly into a vector $\boldsymbol{w} > \boldsymbol{0}$ with rational coordinates, still satisfying $\boldsymbol{w} \cdot \boldsymbol{h} < 0$ for all $\boldsymbol{h} \in H$. Multiplying $\boldsymbol{w}$ by a suitable integer gives a vector $\boldsymbol{v}$ with positive integer coordinates satisfying $\boldsymbol{v} \cdot \boldsymbol{h} < 0$ for all $\boldsymbol{h} \in H$.

(iii) ⇒ (iv) Let $\boldsymbol{v} \in \mathbb{N}^d$, $\boldsymbol{v} > \boldsymbol{0}$, be such that $\boldsymbol{v} \cdot \boldsymbol{h} < 0$ for all $\boldsymbol{h} \in H$. Let L be any linear ordering of $\mathbb{N}^d$ (*e.g.*, the lexicographic one). Define a new linear ordering $L_{\boldsymbol{v}}$ of $\mathbb{N}^d$ by

$$\boldsymbol{p}\, L_{\boldsymbol{v}}\, \boldsymbol{q} \quad \Longleftrightarrow \quad \boldsymbol{v} \cdot \boldsymbol{p} < \boldsymbol{v} \cdot \boldsymbol{q} \ \text{ or } \ (\boldsymbol{v} \cdot \boldsymbol{p} = \boldsymbol{v} \cdot \boldsymbol{q} \ \text{ and } \ \boldsymbol{p}\, L\, \boldsymbol{q}).$$

Note that the equation $\boldsymbol{v} \cdot \boldsymbol{x} = k$ has only finitely many solutions $\boldsymbol{x} \in \mathbb{N}^d$ for any $k \in \mathbb{N}$, because it implies that $0 \leq x_i \leq k/v_i$ for all i. Therefore $L_{\boldsymbol{v}}$ is of order type ω, and since $\boldsymbol{p} \prec_H^+ \boldsymbol{q}$ implies $\boldsymbol{v} \cdot \boldsymbol{p} < \boldsymbol{v} \cdot \boldsymbol{q}$, the dependency relation $\prec_H^+$ can be embedded into $L_{\boldsymbol{v}}$.

(iv) ⇒ (i) Any ordering of type ω is a well-ordering, therefore any transitive relation embeddable into it is well-founded. ■

Additional assertions equivalent to those in Theorem 3 are given in [19, Sec. 3.3, Cor. 2].

Now it is easy to state and prove an existence and uniqueness theorem for recurrences of the form (1) – (2).

Theorem 5 *Let $H \subseteq \mathbb{Z}^d$ be a nonempty set which satisfies any of the equivalent conditions of Theorem* 3. *Then there exists a unique d-dimensional sequence $a : \mathbb{N}^d \to A$ which satisfies* (1) – (2).

Proof: Write $H = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_k\}$. Theorem 3 implies the existence of a well-ordering $<_H$ of $\mathbb{N}^d$ of order type ω which extends the dependency relation $\prec_H^+$. Let $\boldsymbol{p} : \mathbb{N} \to \mathbb{N}^d$ be a bijection satisfying

$$i < j \Longleftrightarrow \boldsymbol{p}_i <_H \boldsymbol{p}_j. \tag{3}$$

The sequence $(a_{\boldsymbol{n}})$ satisfies (1) and (2) if and only if the sequence $f : \mathbb{N} \to A$ defined by $f(i) = a_{\boldsymbol{p}_i}$ satisfies

$$f(i) = \begin{cases} \Phi(f \circ \boldsymbol{p}^{-1}(\boldsymbol{p}_i + \boldsymbol{h}_1), f \circ \boldsymbol{p}^{-1}(\boldsymbol{p}_i + \boldsymbol{h}_2), \ldots, f \circ \boldsymbol{p}^{-1}(\boldsymbol{p}_i + \boldsymbol{h}_k)) & \text{if } \boldsymbol{p}_i \geq \boldsymbol{s}, \\ \varphi(\boldsymbol{p}_i) & \text{if } \boldsymbol{p}_i \not\geq \boldsymbol{s}. \end{cases} \tag{4}$$

We are going to prove, by induction on i, that (4) defines a unique sequence $f(i)$. The *unique* solution of (1) – (2) can then be recovered from f using $a_{\boldsymbol{n}} = f \circ \boldsymbol{p}^{-1}(\boldsymbol{n})$.

Step 0. Let us show that Eqn. (4) completely defines $f(0)$, by proving *ab absurdo* that $\boldsymbol{p}_0 \not\geq \boldsymbol{s}$. If $\boldsymbol{p}_0 \geq \boldsymbol{s}$, then $\boldsymbol{p}_0 + H \subseteq \mathbb{N}^d$ (because $\boldsymbol{s} + H \subseteq \mathbb{N}^d$ by assumption). In other words, $\boldsymbol{p}_0 + \boldsymbol{h}_j \prec_H \boldsymbol{p}_0$ and hence $\boldsymbol{p}_0 + \boldsymbol{h}_j <_H \boldsymbol{p}_0$ for $1 \leq j \leq k$. As H is nonempty and $\mathbf{0} \notin H$, this contradicts Property (3). Thus $\boldsymbol{p}_0 \not\geq \boldsymbol{s}$ and Eqn. (4) determines $f(0) = \varphi(\boldsymbol{p}_0)$.

Step $\boldsymbol{i}$, $\boldsymbol{i} > \mathbf{0}$. Assume $f(0), \ldots, f(i-1)$ are determined uniquely by (4). If $\boldsymbol{p}_i \not\geq \boldsymbol{s}$, then (4) forces $f(i) = \varphi(\boldsymbol{p}_i)$. Otherwise, for $1 \leq j \leq k$, we have $\boldsymbol{p}_i + \boldsymbol{h}_j \prec_H \boldsymbol{p}_i$, hence $\boldsymbol{p}_i + \boldsymbol{h}_j <_H \boldsymbol{p}_i$ and by (3), $\boldsymbol{p}^{-1}(\boldsymbol{p}_i + \boldsymbol{h}_j) < i$. This means that the values of $f \circ \boldsymbol{p}^{-1}(\boldsymbol{p}_i + \boldsymbol{h}_j)$ have already been computed, and $f(i)$ is then uniquely determined by the first equation of (4). ■

This theorem generalizes the result of [23] where $d = 2$ and it is assumed that all $\boldsymbol{h} \in H$ satisfy $h_2 < 0$, or $h_2 = 0$ and $h_1 < 0$.

Example 1 : The knight's walk

We study walks on the lattice $\mathbb{N}^2$ that start anywhere on the lines $x = 0$, $x = 1$, $y = 0$, or $y = 1$, take only two kinds of steps: $(-1, 2)$ and $(2, -1)$, and remain in the region $x \geq 2$, $y \geq 2$ once they have left their starting point. For $m, n \geq 0$, let $a_{m,n}$ denote the number of such walks ending at (m, n). We have

$$a_{m,n} = \begin{cases} a_{m+1,n-2} + a_{m-2,n+1} & \text{if } m, n \geq 2, \\ 1 & \text{otherwise.} \end{cases} \tag{5}$$

This recurrence is of the form (1) – (2), with $H = \{(1, -2), (-2, 1)\}$ and $\boldsymbol{s} = (2, 2)$. The conditions of Theorem 3 are satisfied: for example, if we take $\boldsymbol{v} = (1, 1)$ we have $\boldsymbol{v} \cdot \boldsymbol{h} < 0$ for both $\boldsymbol{h} \in H$. Hence the recurrence has a unique solution whose terms can be computed inductively, for instance diagonal by diagonal. The first few values are given in the following array.

$$\begin{array}{cccccccc} n & & & & & & & \\ \uparrow & & & & & & & \\ 1 & 1 & 5 & 7 & \cdot & \cdot & \cdot & \\ 1 & 1 & 3 & 5 & 10 & 14 & \cdot & \\ 1 & 1 & 3 & 4 & 6 & 10 & \cdot & \\ 1 & 1 & 2 & 2 & 4 & 5 & 7 & \\ 1 & 1 & 2 & 2 & 3 & 3 & 5 & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & \to m \end{array}$$

□

Our restriction to sequences defined in the first orthant might seem a serious limitation of applicability of our results. However, other wedge-like domains can be easily mapped onto the first orthant by a linear transformation, after which our approach applies. For example, if $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d \in \mathbb{Z}^d$ are linearly independent vectors and $\mathcal{C}$ is the convex cone spanned by these vectors then an appropriate transformation could be multiplication by the integer matrix

$$M = \det V \cdot V^{-1}$$

where $V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d]$ is the matrix whose columns are the $\boldsymbol{v}_i$'s. This transforms a recurrence equation in $\mathcal{C}$, with the set of shifts $\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_k\}$, into one in the first orthant, with the set of shifts $\{M\boldsymbol{h}_1, \ldots, M\boldsymbol{h}_k\}$. A transformation of this type is used in Example 5 below. A similar (but slightly different) linear transformation is used in Example 4.

3 Linear recurrences with constant coefficients

In the rest of the paper, we focus our attention on multivariate linear recurrences. Let A be a field of characteristic zero. Let H be a finite nonempty subset of $\mathbb{Z}^d$ satisfying the condition

$$\{\boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} \geq \boldsymbol{0}\} \cap \text{ conv } H = \emptyset. \tag{6}$$

We study the recurrence relation

$$a_{\boldsymbol{n}} = \begin{cases} \sum\limits_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} a_{\boldsymbol{n}+\boldsymbol{h}} & \text{for } \boldsymbol{n} \geq \boldsymbol{s}\,, \\ \\ \varphi(\boldsymbol{n}) & \text{for } \boldsymbol{n} \geq \boldsymbol{0} \text{ and } \boldsymbol{n} \not\geq \boldsymbol{s} \end{cases} \tag{7}$$

where $(c_{\boldsymbol{h}})_{\boldsymbol{h}\in H}$ are given nonzero constants from A. The function φ provides the initial conditions, and we assume that

$$\boldsymbol{s} \in \mathbb{N}^d \quad \text{and} \quad \boldsymbol{s} + H \subseteq \mathbb{N}^d. \tag{8}$$

It is natural to ask how restrictive Condition (6) is. The following discussion suggests that the correct answer is: “not at all”. Let G be a finite nonempty subset of $\mathbb{Z}^d$, and let us consider the linear relation

$$\sum_{\boldsymbol{g}\in G} b_{\boldsymbol{g}} a_{\boldsymbol{n}+\boldsymbol{g}} = 0,$$

where the $b_{\boldsymbol{g}}$ are given nonzero constants from A. This relation can be written in $|G|$ different ways in a form reminiscent of (7): For each $\boldsymbol{g} \in G$,

$$a_{\boldsymbol{n}} = -\sum_{\boldsymbol{g}'\in G\setminus\{\boldsymbol{g}\}} \frac{b_{\boldsymbol{g}'}}{b_{\boldsymbol{g}}}\, a_{\boldsymbol{n}+\boldsymbol{g}'-\boldsymbol{g}} = \sum_{\boldsymbol{h}\in H_{\boldsymbol{g}}} c_{\boldsymbol{h}}\, a_{\boldsymbol{n}+\boldsymbol{h}}$$

with $c_{\boldsymbol{h}} = -b_{\boldsymbol{g}'}/b_{\boldsymbol{g}}$ and $H_{\boldsymbol{g}} = \{\boldsymbol{g}' - \boldsymbol{g};\ \boldsymbol{g}' \in G, \boldsymbol{g}' \neq \boldsymbol{g}\}$. The following proposition implies that at least one of these relations will allow us to compute inductively the numbers $a_{\boldsymbol{n}}$, starting from suitable initial conditions.

Proposition 6 *Let $G \subseteq \mathbb{Z}^d$ be a nonempty finite set. There exists a point $\boldsymbol{g} \in G$ such that the set $H_{\boldsymbol{g}} = \{\boldsymbol{g}' - \boldsymbol{g};\ \boldsymbol{g}' \in G, \boldsymbol{g}' \neq \boldsymbol{g}\}$ satisfies the equivalent conditions of Theorem* 3.

Proof: Let $\boldsymbol{g}$ be the largest point in G with respect to the lexicographic ordering of $\mathbb{Z}^d$. That is, for all $\boldsymbol{g}' \neq \boldsymbol{g}$, there exists $i \in \{1, \ldots, d\}$ such that

$$g_1' = g_1, \ldots, g_{i-1}' = g_{i-1}, \quad \text{and} \quad g_i' < g_i. \tag{9}$$

Then $H_{\boldsymbol{g}}$ satisfies the equivalent conditions of Theorem 3. Let us prove for instance the third one. Let

$$M = \max_{\substack{1 \le j \le d \\ \boldsymbol{g}' \in G}} \left| g_j' - g_j \right|.$$

Let $\boldsymbol{v}$ be the vector $((1+M)^d, \ldots, (1+M)^2, (1+M))$. Let $\boldsymbol{g}' \neq \boldsymbol{g}$, and assume (9) holds. Then

$$\boldsymbol{v} \cdot (\boldsymbol{g}' - \boldsymbol{g}) = \sum_{j=i}^{d} v_j (g_j' - g_j) \le -v_i + M \sum_{j=i+1}^{d} v_j = -(1+M).$$

■

Observe that there can be several "good" choices of $\boldsymbol{g}$. Take for instance the linear equation

$$a_{m,n} = a_{m,n+1} + a_{m+1,n}$$

where $H = \{(0,1), (1,0)\}$ and the conditions of Theorem 3 are not satisfied. Both equivalent relations

$$a_{m,n} = a_{m,n-1} - a_{m+1,n-1} \quad \text{and} \quad a_{m,n} = a_{m-1,n} - a_{m-1,n+1}$$

do satisfy these conditions and hence allow an inductive evaluation of the sequence $a_{m,n}$. Note however that by (8), we need $\boldsymbol{s} \ge (0,1)$ in the former case and $\boldsymbol{s} \ge (1,0)$ in the latter. This means that different sets of initial values are required by these two relations: $(a_{m,0})_{m \ge 0}$ by the former, and $(a_{0,n})_{n \ge 0}$ by the latter.

We now prove an analyticity result.

Theorem 7 *Take $A = \mathbb{C}$. Let $(a_{\boldsymbol{n}})$ be the unique solution of (7). If there are constants $m > 0$ and $\boldsymbol{u} \in \mathbb{R}^d$ such that $|\varphi(\boldsymbol{n})| \le m^{\boldsymbol{u} \cdot \boldsymbol{n}}$ for all $\boldsymbol{n} \not\ge \boldsymbol{s}$, then the generating function of $(a_{\boldsymbol{n}})$*

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{n} \ge \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}}$$

is analytic in a neighborhood of the origin.

Proof: By Theorem 3(iii), there exists $\boldsymbol{v} \in \mathbb{N}^d$ such that $\boldsymbol{v} > \boldsymbol{0}$, and $\boldsymbol{v} \cdot \boldsymbol{h} < 0$ for all $\boldsymbol{h} \in H$. Since H is finite there exists an $\varepsilon > 0$ such that $\boldsymbol{v} \cdot \boldsymbol{h} \le -\varepsilon$ for all $\boldsymbol{h} \in H$. Let

$$M = \max \left\{ 1, \left(\sum_{\boldsymbol{h} \in H} |c_{\boldsymbol{h}}| \right)^{\frac{1}{\varepsilon}}, \max_{1 \le i \le d} m^{\frac{u_i}{v_i}} \right\}.$$

We are going to prove that $|a_{\boldsymbol{n}}| \leq M^{\boldsymbol{v}\cdot\boldsymbol{n}}$ for all $\boldsymbol{n} \in \mathbb{N}^d$, using induction on the well-founded set $(\mathbb{N}^d, \prec_H^+)$.

If $\boldsymbol{n} \not\geq \boldsymbol{s}$ then

$$\begin{aligned} |a_{\boldsymbol{n}}| &= |\varphi(\boldsymbol{n})| \leq m^{\boldsymbol{u}\cdot\boldsymbol{n}} = m^{u_1 n_1} m^{u_2 n_2} \cdots m^{u_d n_d} \\ &= \left(m^{\frac{u_1}{v_1}}\right)^{v_1 n_1} \left(m^{\frac{u_2}{v_2}}\right)^{v_2 n_2} \cdots \left(m^{\frac{u_d}{v_d}}\right)^{v_d n_d} \leq M^{\boldsymbol{v}\cdot\boldsymbol{n}}. \end{aligned}$$

Otherwise we assume inductively that $|a_{\boldsymbol{n}+\boldsymbol{h}}| \leq M^{\boldsymbol{v}\cdot(\boldsymbol{n}+\boldsymbol{h})}$ for all $\boldsymbol{h} \in H$. Then

$$\begin{aligned} \left|\sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} a_{\boldsymbol{n}+\boldsymbol{h}}\right| &\leq \sum_{\boldsymbol{h}\in H} |c_{\boldsymbol{h}}| M^{\boldsymbol{v}\cdot(\boldsymbol{n}+\boldsymbol{h})} \leq \sum_{\boldsymbol{h}\in H} |c_{\boldsymbol{h}}| M^{\boldsymbol{v}\cdot\boldsymbol{n}-\varepsilon} \\ &= M^{\boldsymbol{v}\cdot\boldsymbol{n}-\varepsilon} \sum_{\boldsymbol{h}\in H} |c_{\boldsymbol{h}}| \leq M^{\boldsymbol{v}\cdot\boldsymbol{n}}, \end{aligned}$$

proving the claim. It follows that $F(x_1, x_2, \ldots, x_d)$ converges when $|x_i| < 1/M^{v_i}$. ∎

4 The nature of the solution

Let a be the unique solution of the recursion (7). In this section we investigate the nature of the generating function

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \sum_{\boldsymbol{n}\geq\boldsymbol{s}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}.$$

Let us first recall a few definitions.

Definition 8 *Let A be a field of characteristic zero. Let $F(\boldsymbol{x}) = \sum_{\boldsymbol{n}\geq\boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}}$ be a formal power series in the variables $x_1, x_2, \ldots, x_d$ with coefficients in A. The series F is said to be*

- rational *if there exist polynomials P and Q in $A[x_1, \ldots, x_d]$ such that*

$$F(\boldsymbol{x}) = \frac{P(\boldsymbol{x})}{Q(\boldsymbol{x})};$$

- algebraic *if there exists a nontrivial polynomial P in $d+1$ variables, with coefficients in A, such that*

$$P(x_1, \ldots, x_d, F(\boldsymbol{x})) = 0.$$

- D-finite *if the partial derivatives $\partial^{\boldsymbol{\alpha}} F/\partial \boldsymbol{x}^{\boldsymbol{\alpha}}$ of F, for $\boldsymbol{\alpha} \in \mathbb{N}^d$, span a finite-dimensional vector space over the field $A(x_1, \ldots, x_d)$; or, equivalently, if for $1 \leq i \leq d$, a nontrivial linear differential equation of the form*

$$P_k(\boldsymbol{x}) \frac{\partial^k F}{\partial x_i^k} + \cdots + P_1(\boldsymbol{x}) \frac{\partial F}{\partial x_i} + P_0(\boldsymbol{x}) F = 0$$

  holds, where the polynomials P_ℓ have their coefficients in A.

The series F is transcendental *if it is not algebraic. The coefficients of a D-finite series are said to be* P-recursive.

For properties of D-finite series, see [22] for univariate series and [15] for the multivariate case. We shall clarify below the connection between our linearly recurrent sequences and P-recursiveness.

4.1 From the recurrence relation to a functional equation

Let us now transform our recurrence relation into a functional equation satisfied by the generating function $F_{\boldsymbol{s}}(\boldsymbol{x})$. Multiplying (7) by $\boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}$ and summing over all $\boldsymbol{n} \geq \boldsymbol{s}$ we obtain

$$\begin{aligned} F_{\boldsymbol{s}}(\boldsymbol{x}) &= \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \sum_{\boldsymbol{n}\geq \boldsymbol{s}} a_{\boldsymbol{n}+\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \sum_{\boldsymbol{n}\geq \boldsymbol{s}+\boldsymbol{h}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \\ &= \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \left[F_{\boldsymbol{s}}(\boldsymbol{x}) + P_{\boldsymbol{h}}(\boldsymbol{x}) - M_{\boldsymbol{h}}(\boldsymbol{x})\right] \end{aligned} \tag{10}$$

where

$$P_{\boldsymbol{h}}(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{n}\ngeq \boldsymbol{s} \\ \boldsymbol{n}\geq \boldsymbol{s}+\boldsymbol{h}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} = \sum_{\substack{\boldsymbol{n}\ngeq \boldsymbol{s} \\ \boldsymbol{n}\geq \boldsymbol{s}+\boldsymbol{h}}} \varphi(\boldsymbol{n}) \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \quad \text{and} \quad M_{\boldsymbol{h}}(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{n}\geq \boldsymbol{s} \\ \boldsymbol{n}\ngeq \boldsymbol{s}+\boldsymbol{h}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}. \tag{11}$$

Note the simple relationship between $F_{\boldsymbol{s}}(\boldsymbol{x})$ and the full generating function $F(\boldsymbol{x})$:

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{n}\geq \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} = \boldsymbol{x}^{\boldsymbol{s}} \left(\sum_{\boldsymbol{n}\geq \boldsymbol{s}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} + \sum_{\substack{\boldsymbol{n}\ngeq \boldsymbol{s} \\ \boldsymbol{n}\geq \boldsymbol{0}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \right) = \boldsymbol{x}^{\boldsymbol{s}} (F_{\boldsymbol{s}}(\boldsymbol{x}) + P_{-\boldsymbol{s}}(\boldsymbol{x})).$$

Now rewrite (10) in the form

$$\left(1 - \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}}\right) F_{\boldsymbol{s}}(\boldsymbol{x}) = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \left[P_{\boldsymbol{h}}(\boldsymbol{x}) - M_{\boldsymbol{h}}(\boldsymbol{x})\right]. \tag{12}$$

To clear denominators on the left side of (12) we introduce the notion of *apex* which, as we will show shortly, is related to the nature of the generating function.

Definition 9 *Let $H \subseteq \mathbb{Z}^d$ be a finite set. The* apex *of H is the point $\boldsymbol{p} = (p_1, p_2, \ldots, p_d) \in \mathbb{N}^d$ defined by*

$$p_i = \max\{h_i \; : \; \boldsymbol{h} \in H \cup \{\boldsymbol{0}\}\} \quad (i = 1, 2, \ldots, d).$$

In dimension 2, the apex of H is the upper right corner of the smallest rectangle (with its sides parallel to the axes) enclosing the set $H \cup \{\boldsymbol{0}\}$.

Multiplying (12) by $\boldsymbol{x}^{\boldsymbol{p}}$ where $\boldsymbol{p}$ is the apex of H we obtain

$$Q(\boldsymbol{x}) F_{\boldsymbol{s}}(\boldsymbol{x}) = K(\boldsymbol{x}) - U(\boldsymbol{x}) \tag{13}$$

where

$$Q(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{p}} - \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}}, \tag{14}$$

$$K(\boldsymbol{x}) = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} P_{\boldsymbol{h}}(\boldsymbol{x}), \tag{15}$$

$$U(\boldsymbol{x}) = \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} M_{\boldsymbol{h}}(\boldsymbol{x}), \tag{16}$$

the series $P_{\boldsymbol{h}}$ and $M_{\boldsymbol{h}}$ being given by (11).

From the definition of the apex it follows that $Q(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x}$ called the *characteristic polynomial* or the *kernel* of the recursion. Note that the coefficients of $Q(\boldsymbol{x})$ and $K(\boldsymbol{x})$ are given directly by the coefficients of the recurrence relation and by the initial conditions, respectively. The coefficients of $U(\boldsymbol{x})$ can of course be computed from (7) but are not given explicitly. Therefore we call $K(\boldsymbol{x})$ the *known initial function* and $U(\boldsymbol{x})$ the *unknown initial function*.

The functional equation (13) has a striking feature: on one hand, it is completely equivalent to our recurrence relation, and thus defines uniquely the numbers $a_{\boldsymbol{n}}$ for $\boldsymbol{n} \geq \boldsymbol{s}$, and hence $F_{\boldsymbol{s}}(\boldsymbol{x})$. On the other hand, there seem to be not one, but two unknown functions in it: $F_{\boldsymbol{s}}$ and U. We shall show below on examples how to work with such apparently ambiguous functional equations. If $U(\boldsymbol{x})$ can be found explicitly then the generating function of the unique solution to (7) is given by

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \frac{K(\boldsymbol{x}) - U(\boldsymbol{x})}{Q(\boldsymbol{x})}. \tag{17}$$

Example 1 (continued): The knight's walk

Let us go back to the recurrence (5). We have $\boldsymbol{s} = (2, 2)$, $H = \{(1, -2), (-2, 1)\}$ and the apex is $\boldsymbol{p} = (1, 1)$. Using (11) – (17) we find

$$F_{\boldsymbol{s}}(x, y) = \sum_{m,n \geq 2} a_{m,n} x^{m-2} y^{n-2} = \frac{K(x, y) - U(x, y)}{Q(x, y)}$$

where

$$Q(x, y) = xy - x^3 - y^3,$$

the initial functions being

$$\begin{aligned} K(x, y) &= y^3 \sum_{m \geq 3} \left(a_{m,0} x^{m-2} y^{-2} + a_{m,1} x^{m-2} y^{-1}\right) + x^3 \sum_{n \geq 3} \left(a_{0,n} x^{-2} y^{n-2} + a_{1,n} x^{-1} y^{n-2}\right) \\ &= xy \left(\frac{1+y}{1-x} + \frac{1+x}{1-y}\right), \\ U(x, y) &= \sum_{n \geq 2} a_{2,n} y^{n+1} + \sum_{m \geq 2} a_{m,2} x^{m+1}. \end{aligned}$$

□

The above example shows a strong connection between the nature of the series $F_{\boldsymbol{s}}$ and the nature of the unknown initial function U. We are now going to discuss this connection in full generality. We first need to define the *sections* of a formal power series.

Definition 10 *Let $F(x_1, \ldots, x_d) = \sum_{\boldsymbol{n} \geq \boldsymbol{0}} a_{n_1,\ldots,n_d} \boldsymbol{x}^{\boldsymbol{n}}$ be a formal power series in d variables. A* section *of F is any "sub-series" of F obtained by fixing some of the indices n_i. For instance,*

$$\sum_{n_2,\ldots,n_d \geq 0} a_{1998,n_2,\ldots,n_d} x_2^{n_2} \cdots x_d^{n_d},$$

$$\sum_{n_2,n_4,\ldots,n_d\geq 0} a_{14,n_2,4,n_4,\ldots,n_d} x_2^{n_2} x_4^{n_4}\cdots x_d^{n_d},$$

are sections of F, as well as F itself and $a_{2,4,6,\ldots,2d}$.

This terminology is due to Lipshitz. It is not difficult to prove that all sections of a rational (resp. algebraic, D-finite) series are also rational (resp. algebraic, D-finite). We refer to [15] for a proof in the D-finite case.

Proposition 11 *Let $F_{\boldsymbol{s}}(\boldsymbol{x})$ be the generating function of the unique solution of* (7)*. Then the series $F_{\boldsymbol{s}}(\boldsymbol{x})$ is rational (resp. algebraic, D-finite) if and only if both its known and unknown initial functions $K(\boldsymbol{x})$ and $U(\boldsymbol{x})$ are rational (resp. algebraic, D-finite).*

Proof: If K and U are rational (resp. algebraic, D-finite), then (17) shows that $F_{\boldsymbol{s}}$ is also rational (resp. algebraic, D-finite): the three families of power series under consideration are closed under the sum and the product, and contain rational functions.

Conversely, observe that, for $\boldsymbol{h}\in H$, the series $M_{\boldsymbol{h}}$, defined by (11), and consequently the series U, are finite linear combinations of sections of $F_{\boldsymbol{s}}$. Hence if $F_{\boldsymbol{s}}$ is rational (resp. algebraic, D-finite), then so is $U(\boldsymbol{x})$. Eqn. (17) implies that the same holds for $K(\boldsymbol{x})$. ∎

Note. The series $K(\boldsymbol{x})$ is a linear combination of sections of the full generating function $F(\boldsymbol{x})=\sum_{\boldsymbol{n}\geq\boldsymbol{0}} a_{\boldsymbol{n}}\boldsymbol{x}^{\boldsymbol{n}}$, but not of sections of $F_{\boldsymbol{s}}$.

The above proposition tells us that determining the nature of $F_{\boldsymbol{s}}$ boils down to determining the nature of U (the nature of K is perfectly controlled because K is explicitly given). We give below examples of recurrence relations leading to rational, algebraic irrational, D-finite transcendental, and finally non-D-finite generating functions.

4.2 Rational solutions

Theorem 12 *Assume the apex $\boldsymbol{p}$ of H is $\mathbf{0}$. Then the generating function $F_{\boldsymbol{s}}(\boldsymbol{x})$ of the unique solution of* (7) *is rational if and only if the known initial function $K(\boldsymbol{x})$ itself is rational.*

Proof: For each $\boldsymbol{h}\in H$ we have $\boldsymbol{h}\leq\mathbf{0}$, hence $\boldsymbol{s}+\boldsymbol{h}\leq\boldsymbol{s}$ and $M_{\boldsymbol{h}}(\boldsymbol{x})=\mathbf{0}$. Therefore $U(\boldsymbol{x})=0$ and by (17),

$$F_{\boldsymbol{s}}(\boldsymbol{x})=\frac{K(\boldsymbol{x})}{Q(\boldsymbol{x})}. \tag{18}$$

As $Q(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x}$, it follows that $F_{\boldsymbol{s}}(\boldsymbol{x})$ is a rational function of $\boldsymbol{x}$ if and only if $K(\boldsymbol{x})$ is. ∎

Observe that when $d=1$, we always have rational initial conditions and $\boldsymbol{p}=\mathbf{0}$.

Note that any recurrence with constant coefficients – no matter what the apex – yields a rational generating function under special initial conditions. Assume A is algebraically closed, and take any $\boldsymbol{u}\in A^d$ such that $\sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}}\boldsymbol{u}^{\boldsymbol{h}}=1$. Such a $\boldsymbol{u}$ always exists because H contains a nonzero point. Then $a_{\boldsymbol{n}}=\boldsymbol{u}^{\boldsymbol{n}}$ satisfies $a_{\boldsymbol{n}}=\sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} a_{\boldsymbol{n}+\boldsymbol{h}}$ and the generating function $F(\boldsymbol{x})=\sum_{\boldsymbol{n}\geq\mathbf{0}}\boldsymbol{u}^{\boldsymbol{n}}\boldsymbol{x}^{\boldsymbol{n}}=\prod_{i=1}^{d}(1-u_i x_i)^{-1}$ is rational, as well as $F_{\boldsymbol{s}}$.

4.3 Algebraic solutions

Theorem 13 *Take $A = \mathbb{C}$ and assume that the apex $\boldsymbol{p}$ of H has at most one positive coordinate. Then the generating function $F_{\boldsymbol{s}}(\boldsymbol{x})$ of the unique solution of* (7) *is algebraic if and only if the known initial function $K(\boldsymbol{x})$ itself is algebraic.*

Before we prove this theorem, let us first study a particular recurrence, which will illustrate the main ideas of the proof.

Example 2: Dyck paths

For $m, n \geq 0$, let $a_{m,n}$ denote the number of paths on $\mathbb{N}^2$ that start from the origin $(0,0)$, take only two kinds of steps: $(1,1)$ and $(1,-1)$, and never touch the horizontal axis once they have left the origin. We are especially interested in paths that end on the line $y = 1$: we call them *Dyck paths.* We have

$$a_{m,n} = \begin{cases} a_{m-1,n-1} + a_{m-1,n+1} & \text{if } m, n \geq 1, \\ \delta_{(m,n),(0,0)} & \text{if } m = 0 \text{ or } n = 0. \end{cases}$$

This is a linear recurrence with constant coefficients, with $H = \{(-1,-1), (-1,1)\}$ and $\boldsymbol{s} = (1,1)$. The apex is $\boldsymbol{p} = (0,1)$. The first values of $a_{m,n}$ are given below.

$$\begin{array}{cccccccccccl}
n \\
\uparrow \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 6 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 5 & 0 & 20 \\
0 & 0 & 0 & 0 & 1 & 0 & 4 & 0 & 14 & 0 \\
0 & 0 & 0 & 1 & 0 & 3 & 0 & 9 & 0 & 28 \\
0 & 0 & 1 & 0 & 2 & 0 & 5 & 0 & 14 & 0 \\
0 & 1 & 0 & 1 & 0 & 2 & 0 & 5 & 0 & 14 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rightarrow & m
\end{array}$$

Using (11) – (17) we obtain:

$$F_{\boldsymbol{s}}(x,y) = \sum_{m,n \geq 1} a_{m,n} x^{m-1} y^{n-1} = \frac{K(x,y) - U(x,y)}{Q(x,y)},$$

with

$$Q(x,y) = y - x - xy^2,$$

the initial functions being

$$K(x,y) = y \qquad \text{and} \qquad U(x,y) = \sum_{m \geq 1} a_{m,1} x^m \equiv U(x).$$

In other words,

$$\left(y - x - xy^2\right) F_{\boldsymbol{s}}(x,y) = y - U(x). \tag{19}$$

Observe that $U(x)$ is the length generating function for Dyck paths. We are now going to use an idea that occurs in various places (*e.g.*, [12, Ex. 2.2.1.4 and 2.2.1.11], [4], [17], [7],

[2]) and is sometimes called the *kernel method* [1]: let $\xi(x)$ be the formal power series in x defined by

$$Q(x, \xi(x)) = \xi(x) - x - x\xi(x)^2 = 0.$$

Replacing y by $\xi(x)$ in (19) shows that

$$U(x) = \xi(x) = \frac{1 - \sqrt{1 - 4x^2}}{2x},$$

from which we can express $F_{\boldsymbol{s}}(x, y)$, which is also an algebraic function. $\square$

We are now going to generalize the kernel method into a proof of Theorem 13.

Proof of Theorem 13: According to Proposition 11, if $F_{\boldsymbol{s}}$ is algebraic, then so is K. Let us prove that the converse is also true.

If $\boldsymbol{p} = \mathbf{0}$, then the proof is similar to that of the previous theorem. Assume now that exactly one coordinate of $\boldsymbol{p}$ is positive. Without loss of generality, assume that $p_1 = \cdots = p_{d-1} = 0$ and $p_d > 0$. Then

$$\begin{aligned} U(\boldsymbol{x}) &= \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} \sum_{\substack{\boldsymbol{n} \geq \boldsymbol{s} \\ \boldsymbol{n} \not\geq \boldsymbol{s}+\boldsymbol{h}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \\ &= \sum_{\substack{\boldsymbol{h} \in H \\ h_d > 0}} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} \sum_{n_d = s_d}^{s_d + h_d - 1} \sum_{\substack{(n_1, \ldots, n_{d-1}) \geq \\ (s_1, \ldots, s_{d-1})}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}. \end{aligned}$$

This shows that $U(\boldsymbol{x})$ is a *polynomial* in x_d of degree at most $p_d - 1$. Our functional equation (13) reads

$$U(\boldsymbol{x}) = K(\boldsymbol{x}) - Q(\boldsymbol{x}) F_{\boldsymbol{s}}(\boldsymbol{x}), \tag{20}$$

where

$$Q(\boldsymbol{x}) = x_d^{p_d} - \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}}$$

is a polynomial whose degree in x_d is at least p_d. We shall prove that $Q(\boldsymbol{x})$ regarded as a polynomial in x_d admits (at least) p_d roots $\xi_i(x_1, \ldots, x_{d-1})$, counted with multiplicities, such that

$$\xi_i(0, \ldots, 0) = 0. \tag{21}$$

Assuming for the moment that this is true, replacing x_d by $\xi_i(x_1, \ldots, x_{d-1})$ in (20) tells us that[1], if ξ_i is a root of Q of multiplicity m, then

$$U(\xi_i) = K(\xi_i), \quad U'(\xi_i) = K'(\xi_i), \ldots, \quad U^{(m-1)}(\xi_i) = K^{(m-1)}(\xi_i),$$

the derivatives being taken with respect to x_d. The p_d roots of Q thus provide a total of p_d equations for the polynomial U, of degree at most $p_d - 1$. We can then reconstruct U by means of the Hermite interpolation formula (see, *e.g.*, [10, Sec. 6.1, Problem 10]), or, when

[1] Condition (21) is required to make the substitution of ξ_i for x_d legitimate because as an equation between convergent power series, (20) is valid only in some neighbourhood of the origin.

Q has no double roots, by means of the well-known Lagrange interpolation formula. Because the ξ_i's are algebraic functions of $x_1, \ldots, x_{d-1}$, this shows that $U(\boldsymbol{x})$ is algebraic provided $K(\boldsymbol{x})$ is. The same holds for $F_{\boldsymbol{s}}(\boldsymbol{x})$.

To prove the existence of the roots ξ_i, let us observe that

$$Q(0, \ldots, 0, x_d) \;=\; x_d^{p_d} \;-\; \sum_{\substack{\boldsymbol{h} \in H \\ h_1 = \cdots = h_{d-1} = 0}} c_{\boldsymbol{h}} x_d^{p_d - h_d}.$$

Because $\{\boldsymbol{x} \in \mathbb{R}^d;\; \boldsymbol{x} \geq \boldsymbol{0}\} \;\cap\; H \;=\; \emptyset$, we have $h_d < 0$ for all $\boldsymbol{h} \in H$ such that $h_1 = \cdots = h_{d-1} = 0$. Therefore $p_d - h_d > p_d$ for all such $\boldsymbol{h}$, which implies that $x_d = 0$ is a root of $Q(0, \ldots, 0, x_d)$ of multiplicity p_d. It follows that at least p_d of the roots of Q satisfy (21). ■

Remarks

1. The proof of Theorem 13 not only shows that under the stated conditions the generating functions of the solution are algebraic, but also provides an *algorithm* to compute them. They are given as rational expressions in the roots of the algebraic equation $Q(\boldsymbol{x}) = 0$ and hence belong to an algebraic extension of $\mathbb{C}(\boldsymbol{x})$. If the algebraic equations satisfied by the generating functions themselves are desired, they can be obtained by some routine resultant computations. However, in general the degree of these equations will be higher than the degree of Q in x_d.

2. It happens quite often, in recurrence relations coming from enumerative combinatorics, that the known initial function $K(\boldsymbol{x})$ itself is a polynomial in x_d. In this case, the polynomial $K - U$ has at least p_d roots (namely, $\xi_1, \ldots, \xi_{p_d}$). The polynomial U having degree at most $p_d - 1$ (in x_d), this implies that K has degree at least p_d. If K has *exactly* degree p_d, then

$$K(\boldsymbol{x}) - U(\boldsymbol{x}) = \mathrm{lc}\,(K) \prod_{i=1}^{p_d} (x_d - \xi_i)$$

where $\mathrm{lc}\,(K)$ denotes the leading coefficient of K with respect to x_d, and Eqn. (20) implies that

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \frac{\mathrm{lc}\,(K)}{Q(\boldsymbol{x})} \prod_{i=1}^{p_d} (x_d - \xi_i). \tag{22}$$

The degree of Q in x_d is $p_d + r$, where $r = \max\{-h_d, \boldsymbol{h} \in H \cup \{\boldsymbol{0}\}\}$. Assume Q factors as

$$Q(\boldsymbol{x}) = \mathrm{lc}\,(Q) \prod_{i=1}^{p_d} (x_d - \xi_i) \prod_{i=1}^{r} (x_d - \mu_i).$$

Then we can rewrite the series $F_{\boldsymbol{s}}$ as

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \frac{\mathrm{lc}\,(K)}{\mathrm{lc}\,(Q)} \prod_{i=1}^{r} \frac{1}{x_d - \mu_i}. \tag{23}$$

Eqns. (22) and (23) show that the two cases $p_d = 1$ and $r = 1$, corresponding respectively to $\max h_d = 1$ and $\min h_d = -1$, are especially favourable. In both cases, the series $F_{\boldsymbol{s}}(x)$ is, up to an explicit rational-function factor, equal to $(x_d - \chi)^{\pm 1}$, where χ is a solution of $Q(x_1, \ldots, x_{d-1}, \chi) = 0$. In this case, we can write immediately the algebraic equation satisfied by $F_{\boldsymbol{s}}(\boldsymbol{x})$. Examples will be given below.

Example 3: Generalized Dyck paths

In the problem of *generalized Dyck paths* [14, 13, 5], we are given a finite set of steps $\mathcal{S} = \{(r_1, s_1), \ldots, (r_k, s_k)\}$ where $r_i, s_i \in \mathbb{Z}$ and $r_i > 0$. Let $b_{m,n}$ denote the number of paths going from $(0,0)$ to (m,n), using only steps from $\mathcal{S}$ and staying within the first quadrant. We are mainly interested in the numbers $b_{m,0} \equiv b_m$. Obviously, the numbers $b_{m,n}$ satisfy

$$b_{m,n} = \begin{cases} \displaystyle\sum_{i:\,(r_i,s_i)\le(m,n)} b_{m-r_i,n-s_i} & \text{if } m,n \ge 0 \text{ and } (m,n) \ne (0,0), \\ 1 & \text{if } m = n = 0. \end{cases}$$

This does not fit (7), but can be easily brought into the desired form. Let $r = \max_{1\le i\le k} r_i$ and $s = \max_{1\le i\le k} s_i$. Attach r columns of zeros to the left of the array b, and s rows of zeros below. Call the resulting array a and number its rows and columns starting with 0. Then

$$b_{m,n} = a_{m+r,n+s} \quad \text{for } m,n \ge 0.$$

Let (ρ, σ) be any step in $\mathcal{S}$ which is maximal with respect to the partial order $\le$ (*e.g.*, the lexicographically largest step). If $a_{r-\rho,s-\sigma}$ is reset to 1 then a satisfies

$$a_{m,n} = \begin{cases} a_{m-r_1,n-s_1} + \cdots + a_{m-r_k,n-s_k} & \text{if } m \ge r \text{ and } n \ge s, \\ \delta_{(m,n),(r-\rho,s-\sigma)} & \text{if } m < r \text{ or } n < s. \end{cases}$$

This clearly fits (7) with $H = \{(-r_1, -s_1), \ldots, (-r_k, -s_k)\}$, $\boldsymbol{s} = (r,s)$, and $c_{\boldsymbol{h}} = 1$ for all $\boldsymbol{h} \in H$. From the fact that $r_i > 0$, it follows that $\{\boldsymbol{x} \in \mathbb{R}^d;\ \boldsymbol{x} \ge \boldsymbol{0}\} \cap \operatorname{conv} H = \emptyset$. Hence the conditions of Theorem 3 are satisfied. The apex of H is $\boldsymbol{p} = (0, \max\{0,t\})$ with $t = -\min\{s_1, \ldots, s_k\}$. We distinguish two cases.

• If $s_i \ge 0$ for all i, then the apex is $(0,0)$. We can use formula (18) with $Q(x,y) = 1 - x^{r_1}y^{s_1} - \cdots - x^{r_k}y^{s_k}$ and $K(x,y) = 1$. We obtain the rational (and expected!) generating function

$$G(x,y) = \sum_{m,n\ge0} b_{m,n}x^m y^n = \sum_{\substack{m\ge r\\ n\ge s}} a_{m,n}x^{m-r}y^{n-s} = F_{\boldsymbol{s}}(x,y) = \frac{1}{1 - x^{r_1}y^{s_1} - \cdots - x^{r_k}y^{s_k}}.$$

The generating function for paths ending on the horizontal axis is

$$g(x) = \sum_{m\ge0} b_m x^m = G(x,0) = \frac{1}{1 - \displaystyle\sum_{\substack{1\le i\le k\\ s_i=0}} x^{r_i}}.$$

• If there exists i such that $s_i < 0$, then the apex of H is $(0,t)$ with $t > 0$, and the corresponding generating functions are algebraic. Using (11) – (15) we find

$$Q(x,y) = y^t - x^{r_1}y^{s_1+t} - \cdots - x^{r_k}y^{s_k+t}, \quad K(x,y) = y^t.$$

We are now in the framework described in the second remark that follows the proof of Theorem 13. The polynomial Q has degree $t+s$ in y, with leading coefficient $-\sum_{i:\,s_i=s} x^{r_i}$.

(We assume $s > 0$ to avoid trivial cases.) Let $\xi_1(x), \ldots, \xi_t(x)$ be t solutions of $Q(x, \xi(x)) = 0$ satisfying $\xi(0) = 0$. Let $\mu_1(x), \ldots, \mu_s(x)$ be the s remaining solutions. We find explicit expressions for the desired series:

$$G(x,y) = \sum_{m,n\geq 0} b_{m,n} x^m y^n = F_{\mathbf{s}}(x,y) = \frac{1}{Q(x,y)} \prod_{i=1}^{t} (y - \xi_i(x)) = -\frac{1}{\sum_{i:s_i=s} x^{r_i}} \prod_{i=1}^{s} \frac{1}{y - \mu_i(x)} \tag{24}$$

and

$$g(x) = \sum_{m\geq 0} b_m x^m = G(x,0) = \frac{(-1)^t}{Q(x,0)} \prod_{i=1}^{t} \xi_i(x) = \frac{(-1)^{s+1}}{\sum_{i:s_i=s} x^{r_i}} \prod_{i=1}^{s} \frac{1}{\mu_i(x)}. \tag{25}$$

An alternative solution to this problem is given in [5] in terms of a system of algebraic equations defining $g(x)$.

As indicated in the remark above, two particular cases happen to be especially simple (and combinatorially equivalent, as far as $g(x)$ is concerned). If $t = 1$, *i.e.*, $\min s_i = -1$, we find $g(x) = -\xi_1(x)/Q(x,0)$, so that the algebraic equation satisfied by g is

$$Q\left(x,\ g(x) \textstyle\sum_{i:\, s_i=-1} x^{r_i}\right) = 0.$$

In other words, denoting $= g(x) \sum_{i:\, s_i=-1} x^{r_i}$:

$$\hat{g} - \sum_{i=1}^{k} x^{r_i} \hat{g}^{s_i+1} = 0. \tag{26}$$

If $s = 1$, *i.e.*, $\max s_i = 1$, we find $g(x) = (\mu_1(x) \sum_{i:\, s_i=1} x^{r_i})^{-1}$, so that a similar algebraic equation holds:

$$Q\left(x,\ \frac{1}{g(x) \sum_{i:\, s_i=1} x^{r_i}}\right) = 0.$$

In other words, denoting $\tilde{g}(x) = g(x) \sum_{i:\, s_i=1} x^{r_i}$, and after a multiplication by $\tilde{g}^{t+1}$:

$$\tilde{g} - \sum_{i=1}^{k} x^{r_i} \tilde{g}^{1-s_i} = 0. \tag{27}$$

The transformation $(r_i, s_i) \longrightarrow (r_i, -s_i)$, which boils down to reading the paths from right to left, shows the combinatorial equivalence between the two problems, reflected by Eqns. (26) – (27).

As special cases, this example includes some well-known lattice-path enumeration problems. For instance, these paths are called:

- *Dyck paths*, if $\mathcal{S} = \{(1,-1),(1,1)\}$. In this case, $Q(x,y) = y - xy^2 - x$, and from (25),

$$g(x) = \frac{1 - \sqrt{1 - 4x^2}}{2x^2}.$$

- *Motzkin paths*, if $\mathcal{S} = \{(1,-1),(1,1),(1,0)\}$. In this case, $Q(x,y) = y - xy^2 - x - xy$, and from (25),

$$g(x) = \frac{1 - x - \sqrt{1-2x-3x^2}}{2x^2}.$$

- *Schröder paths*, if $\mathcal{S} = \{(1,-1),(1,1),(2,0)\}$. In this case, $Q(x,y) = y - xy^2 - x - x^2y$, and from (25),

$$g(x) = \frac{1 - x^2 - \sqrt{1-6x^2+x^4}}{2x^2}.$$

In the same way we could count generalized Dyck paths with coloured steps – then the corresponding coefficient $c_{\boldsymbol{h}}$ would equal the number of colours available for step $\boldsymbol{h}$. This approach generalizes without significant change to higher-dimensional paths if the steps have positive coordinates in all but perhaps one (fixed) dimension. In all these cases, the generating functions are algebraic. □

The next example is actually an application of the previous one.

Example 4: Directed paths above a line of rational slope

Let p and q be positive integers. Consider the problem of counting lattice paths with steps $(1,0)$ and $(0,1)$ which start at the origin and stay on or above the line $qy = px$. Using the linear transformation $\mathcal{L} : (i,j) \mapsto (i+j, qj - pi)$ we obtain the equivalent problem of counting lattice paths with steps $(1,-p)$ and $(1,q)$ which start at the origin and stay within the first quadrant. This is because $\mathcal{L}$ maps the "forbidden" region below the line $qy = px$ into the region below the line $y = 0$. We are thus back to Example 3, with $S = \{(1,-p),(1,q)\}$. The apex is $\boldsymbol{p} = (0,p)$ and we are in the algebraic case, with $t = p$ and $s = q$. We have $Q(x,y) = y^p - xy^{p+q} - x$ and the relevant generating functions can be derived from (24) and (25):

$$G(x,y) = \frac{1}{y^p - xy^{p+q} - x} \prod_{i=1}^{p} (y - \xi_i(x)) = -\frac{1}{x} \prod_{i=1}^{q} \frac{1}{y - \mu_i(x)},$$

$$g(x) = \frac{(-1)^{p+1}}{x} \prod_{i=1}^{p} \xi_i(x) = \frac{(-1)^{q+1}}{x} \prod_{i=1}^{q} \frac{1}{\mu_i(x)}.$$

The series G and g can be understood in terms of the original paths (with unit steps) as follows: $G(x,y)$ counts these paths according to their length (variable x) and an additional statistics that describes some kind of "distance" between the endpoint of the path and the line $px = qy$. The series $g(x)$ is the length generating function for paths ending on that line.

The cases $p = 1$ and $q = 1$ are especially simple; moreover, they are equivalent, as far as the evaluation of $g(x)$ is concerned. If, for instance, $p = 1$, then $g(x) = \xi(x)/x$ where $\xi(x)$ is the unique power-series solution of

$$\xi = x\left(1 + \xi^{q+1}\right).$$

The Lagrange inversion formula gives then

$$g(x) = \sum_{n \ge 0} b_n x^n = \sum_{m \ge 0} \frac{x^{m(q+1)}}{1+mq} \binom{m(q+1)}{m}.$$

Note that $b_{m(q+1)}$ equals the number of paths from the origin to the point (mq, m) in our original problem. We have thus recovered a classical result (*e.g.* [6, Thm. 2]).

By changing the initial conditions, we could count the paths that start at the origin and stay above the line $px = qy$ (if $q = 1$, then this question is equivalent to the initial one; for the case $p = 1$, see [9, Theorem 5.5]). Instead of the steps $(1, 0)$ and $(0, 1)$ we could take any set of steps with nonnegative components, and still obtain algebraic generating functions. □

4.4 D-finite transcendental solutions

According to the two previous subsections, the solution of (7) can only be transcendental if the apex $\boldsymbol{p}$ has more than one positive coordinate (unless of course the known initial function $K(\boldsymbol{x})$ is already transcendental).

Example 5: Ballot problems, Young tableaux and involutions avoiding long increasing subsequences

In an election, d candidates $C_1, \ldots, C_d$ receive respectively $m_1, \ldots, m_d$ votes, with $m_1 \geq \cdots \geq m_d \geq 0$. The problem is to determine the number $b_{\boldsymbol{m}}$ of ways one can count the votes so that at any stage, C_i has at least as many votes as C_{i+1}, for $1 \leq i < d$.

Clearly, such counting processes can be encoded by paths on the lattice $\mathbb{N}^d$ that start from the origin $\mathbf{0}$, end at $\boldsymbol{m}$, take unit steps in the d positive directions, and have all their vertices in the wedge $x_1 \geq x_2 \geq \cdots \geq x_d \geq 0$.

These paths are known to encode Young tableaux of height at most d: a step of the path going in the ith direction means that an entry is added to the ith row of the tableau (Figure 1). Equivalently, via the Robinson-Schensted correspondence, these paths encode involutions of the symmetric group $\mathcal{S}_n$ that avoid the pattern $123\ldots(d+1)$, *i.e.*, have no increasing sub-sequence of length $> d$.

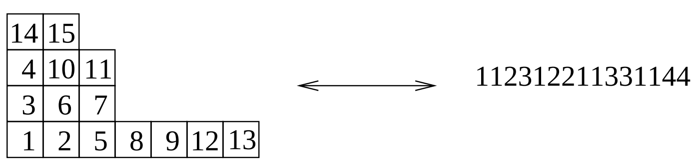


Figure 1: A Young tableau of height 4 and the word – or path – that encodes it (a letter i in the word indicates that the path takes a step in the ith direction).

The numbers $b_{\boldsymbol{m}}$, for $\boldsymbol{m} \in \mathbb{Z}^d$, satisfy:

$$b_{\boldsymbol{m}} = \begin{cases} \sum_{i=1}^{d} b_{\boldsymbol{m}-\boldsymbol{e}_i} & \text{if } m_1 \geq \cdots \geq m_d \geq 0 \text{ and } \boldsymbol{m} \neq \mathbf{0}, \\ \delta_{\boldsymbol{m},\mathbf{0}} & \text{otherwise} \end{cases}$$

where $\boldsymbol{e}_i$ denotes the unit vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 in ith position. We can transform this recurrence into a recurrence of type (7) by setting

$$a_{\boldsymbol{n}} = b_{n_1+n_2+\cdots+n_d-d, \ldots, n_{d-1}+n_d-2, n_d-1},$$

except for the value $a_{0,1,\ldots,1}$ which we find convenient to reset to 1. Then for $\boldsymbol{n} \geq \boldsymbol{0}$,

$$a_{\boldsymbol{n}} = \begin{cases} \sum_{\boldsymbol{h}\in H} a_{\boldsymbol{n}+\boldsymbol{h}} & \text{if } \boldsymbol{n} \geq \boldsymbol{1}, \\ \delta_{\boldsymbol{n},(0,1,\ldots,1)} & \text{otherwise} \end{cases}$$

where

$$H = \{(-1,0,\ldots,0),(1,-1,0,\ldots,0),(0,1,-1,0,\ldots,0),\ldots,(0,\ldots,0,1,-1)\}.$$

The set H satisfies Theorem 3(iii) with $\boldsymbol{v} = (1,2,\ldots,d)$, and thus the recurrence defines a unique sequence. The apex is $(1,1,\ldots,1,0)$.

Proposition 14 *The generating function $F_{\boldsymbol{1}}(\boldsymbol{x}) = \sum_{\boldsymbol{n}\geq\boldsymbol{1}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{1}}$ is D-finite for all d. It is rational if $d=1$, algebraic of degree 2 (quadratic) if $d=2$, and transcendental if $d \geq 3$.*

Proof: This ballot problem was solved by MacMahon [18, Sec.III, Chap.V, 103]:

$$b_{\boldsymbol{m}} = \prod_{1\leq i<j\leq d} (m_i - m_j + j - i) \frac{\left(\sum_{i=1}^d m_i\right)!}{\prod_{i=1}^d (m_i + d - i)!}.$$

This gives, for $\boldsymbol{n} \geq \boldsymbol{1}$,

$$a_{\boldsymbol{n}} = \prod_{1\leq i<j\leq d} (n_i + \cdots + n_{j-1}) \frac{\left(\sum_{i=1}^d i n_i - \binom{d+1}{2}\right)!}{\prod_{i=1}^d (n_i + \cdots + n_d - 1)!}.$$

We prove that this d-dimensional sequence is P-recursive by combining the P-recursiveness of some elementary sequences with the fact that the class of P-recursive sequences is closed under the Hadamard product (see [15] for details). Hence $F_{\boldsymbol{1}}$ is D-finite.

Now, if $F_{\boldsymbol{1}}(\boldsymbol{x})$ is algebraic, then so is its section

$$f(z) = \sum_{n\geq 1} a_{1,\ldots,1,n}\, z^n = \sum_{n\geq 1} b_{n-1,\ldots,n-1}\, z^n,$$

corresponding to Young tableaux of rectangular shape. But

$$a_{1,\ldots,1,n+1} = \prod_{i=1}^{d-1} i^{d-i} \frac{(dn)!}{\prod_{i=0}^{d-1}(n+i)!} \sim \prod_{i=1}^{d-1} i^{d-i} \frac{\sqrt{d}}{(2\pi)^{(d-1)/2}} \cdot \frac{d^{dn}}{n^{(d^2-1)/2}}$$

by Stirling's formula. For $f(z)$ to be algebraic, it is necessary (see [8, Thm. D]) that

$$\frac{d^2-1}{2} \notin \{1,2,3,\ldots\},$$

which rules out all odd values of d, except $d=1$, and that

$$\prod_{i=1}^{d-1} i^{d-i} \frac{\sqrt{d}}{(2\pi)^{(d-1)/2}} \cdot \Gamma\left(\frac{3-d^2}{2}\right)$$

be an algebraic number. If d is even, then $\Gamma\left(\frac{3-d^2}{2}\right)$ is, up to a rational factor, equal to $\Gamma(1/2) = \sqrt{\pi}$. The above condition reads "$\pi^{d/2-1}$ is algebraic", and forces $d = 2$. Thus $F_{\mathbf{1}}$ is transcendental when $d \geq 3$.

If $d = 1$, then $a_{n_1} = 1$ for $n_1 \geq 1$. Hence $F_1 = 1/(1 - x_1)$ is rational.

If $d = 2$, then the apex is $(1, 0)$ and the known initial function is rational. We apply the technique of Section 4.3 to obtain

$$F_{\mathbf{1}}(x_1, x_2) = \frac{1}{2}\,\frac{2x_1 - 1 + \sqrt{1 - 4x_2}}{x_1 - x_1^2 - x_2}.$$

Note that this calculation is completely equivalent to that of Example 2, as ballot paths are merely left factors of Dyck paths.

■

4.5 Non-D-finite (hypertranscendental) solutions

Our last example is a variation on the knight's walk (Example 1). It shows that the solution of (7) need not be D-finite, even though the initial conditions are rational.

Example 6 Let $(a_{m,n})_{m,n\geq 0}$ be the sequence defined by:

$$a_{m,n} = \begin{cases} a_{m+1,n-2} + a_{m-2,n+1} - a_{m-1,n-1} & \text{if } m, n \geq 2, \\ -\delta_{(m,n),(1,1)} & \text{if } m \leq 1 \text{ or } n \leq 1. \end{cases}$$

The first few values of $a_{m,n}$ are given below.

$$\begin{array}{llllllll}
n & & & & & & & \\
\uparrow & & & & & & & \\
0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \\
0 & 0 & -1 & 0 & 0 & 1 & \cdot & \\
0 & 0 & 0 & 0 & -1 & 0 & \cdot & \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & \\
0 & 0 & 1 & 0 & 0 & -1 & 0 & \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \to\ m
\end{array}$$

This is a problem of type (7) with $H = \{(1, -2), (-2, 1), (-1, -1)\}$, $\boldsymbol{s} = (2, 2)$, and $\boldsymbol{p} = (1, 1)$. We find, using (11) – (17)

$$Q(x, y) = (x - y^2)(y - x^2), \quad K(x, y) = xy, \quad \text{and} \quad U(x, y) = G(x) + G(y)$$

with

$$G(x) = \sum_{m\geq 2} a_{m,2}x^{m+1}.$$

We have used the symmetry of the problem in m and n. As usual, let $F_s(x,y) = \sum_{m,n\geq 2} a_{m,n} x^{m-2} y^{n-2}$. Then we have from (13)

$$\left[x - y^2\right]\left[y - x^2\right] F_s(x,y) = xy - G(x) - G(y). \tag{28}$$

Replacing y by x^2 in (28), we obtain the following equation defining G:

$$x^3 - G(x) - G\left(x^2\right) = 0,$$

Iterating this equation shows that, for $n \geq 1$,

$$G(x) = \sum_{i=0}^{n-1} (-1)^i x^{3\cdot 2^i} + (-1)^n G\left(x^{2^n}\right).$$

By definition, $G(x) = O(x^3)$, and therefore

$$G(x) = \sum_{i\geq 0} (-1)^i x^{3\cdot 2^i}.$$

The series G is a *lacunary series*; in particular, its coefficients c_n do not satisfy any non-trivial recurrence relation of the type

$$P_0(n)c_n + P_1(n)c_{n-1} + \cdots + P_k(n)c_{n-k} = 0,$$

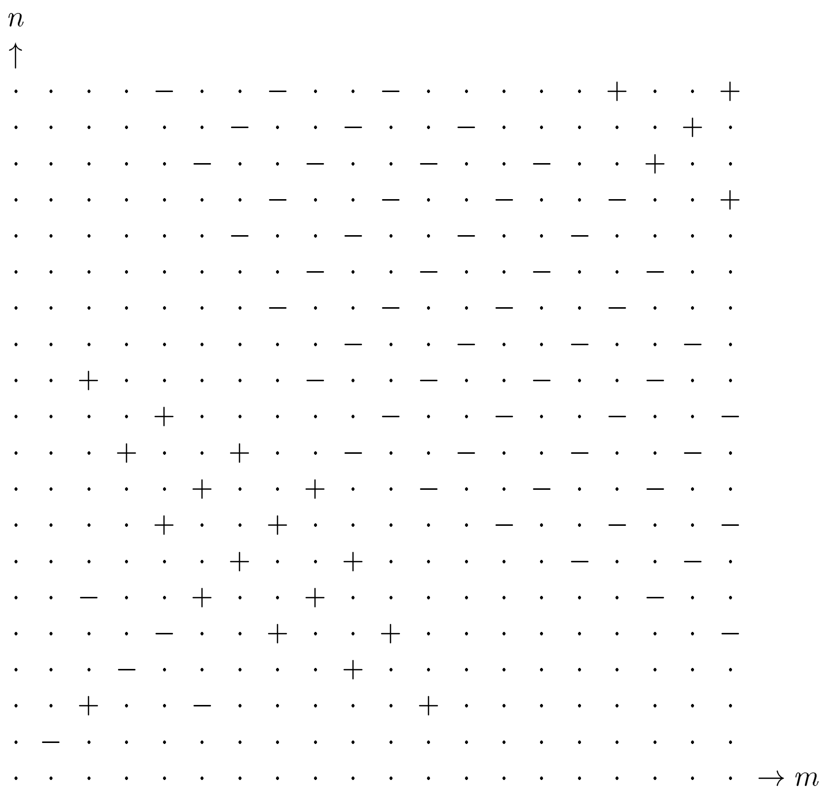
where the $P_i(n)$ are polynomials in n. Such recurrence relations characterize D-finite series: hence $G(x)$ cannot be D-finite. Consequently, the series $F_s(x,y)$ itself is not D-finite.

We can actually state a stronger result: according to [16, Thm. 4], the series G is *hypertranscendental*, meaning that it does not satisfy *any algebraic differential equation* of the form

$$P(x, G(x), G'(x), G''(x), \ldots, G^{(k)}(x)) = 0$$

where P is a polynomial with complex coefficients of $k+2$ variables.

The numbers $a_{m,n}$ actually belong to $\{0, -1, 1\}$ and we could write down an explicit expression for them. But the following table, where the zero entries are replaced by dots and the nonzero entries are replaced by their signs, shows the underlying lozenge pattern and makes a long formula short.



$\square$

Example 1 (continued): what about the knight?
Let us recall the functional equation we have obtained for this problem:

$$[xy - x^3 - y^3]F_{\boldsymbol{s}}(x, y) = K(x, y) - G(x) - G(y) \tag{29}$$

where

$$K(x, y) = xy\left(\frac{1+y}{1-x} + \frac{1+x}{1-y}\right) \qquad \text{and} \qquad G(x) = \sum_{m\geq 2} a_{m,2}x^{m+1}.$$

We have used the symmetry of the problem in m and n. We are going to show that the kernel method allows once again to solve completely this functional equation. Let $\xi(x) = O(x^2)$ be the formal power series in x satisfying

$$x\xi - x^3 - \xi^3 = 0.$$

The Lagrange inversion formula, applied to $\xi(x)/x$, provides an explicit expression:

$$\xi(x) = x^2 \sum_{m\geq 0} \frac{x^{3m}}{2m+1}\binom{3m}{m} = O(x^2).$$

Replacing y by $\xi(x)$ in (29) gives a functional equation that defines $G(x)$:

$$K(x, \xi(x)) - G(x) - G(\xi(x)) = 0.$$

As $G(x) = O(x^2)$, we obtain, after iterating this equation infinitely many times

$$G(x) = \sum_{i \geq 0} (-1)^i K\left[\xi^{(i)}(x), \xi^{(i+1)}(x)\right]$$

where $\xi^{(i)} = \xi \circ \cdots \circ \xi$ is the ith iterate of ξ. Replacing $G(x)$ by the above explicit value in (29) gives an expression for $F_{\boldsymbol{s}}$. It has been proved that G is irrational [20]. It is in fact not D-finite [3]. □

Final remarks

1. The approach of this paper can be generalized without any significant alteration to recurrence relations involving an inhomogeneous term:

$$a_{\boldsymbol{n}} = \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} a_{\boldsymbol{n}+\boldsymbol{h}} + b_{\boldsymbol{n}},$$

where $(b_{\boldsymbol{n}})_{\boldsymbol{n} \geq \boldsymbol{s}}$ is a given d-dimensional sequence in the field A. The inhomogeneous term only affects the known initial function $K(\boldsymbol{x})$, which becomes

$$K(\boldsymbol{x}) = \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} x^{\boldsymbol{p}-\boldsymbol{h}} P_{\boldsymbol{h}}(\boldsymbol{x}) + B(\boldsymbol{x}), \qquad \text{with} \qquad B(\boldsymbol{x}) = \sum_{\boldsymbol{n} \geq \boldsymbol{s}} b_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}.$$

2. Some of the results and techniques presented in this paper do not require H to be a finite set. If H is infinite, the initial conditions have to include $a_{\boldsymbol{n}} = 0$ for $\boldsymbol{n} \not\geq \mathbf{0}$.

3. We used the kernel method to prove Proposition 13 and solve Examples 6 and 1. More examples suggest that the roots of the kernel will *always* provide enough information on the unknown initial function $U(\boldsymbol{x})$ to characterize it completely, thus yielding some kind of solution of the recursion; this could, at least in the two-dimensional case, allow for a complete classification of the recursions according to the nature of the associated generating function.

Acknowledgements. The authors wish to express their thanks to Ira Gessel for providing the basis of Example 5, as well as several useful references.

References

[1] C. Banderier, *Combinatoire analytique : application aux marches aléatoires*, Mémoire de DEA, INRIA Rocquencourt, June 1998.

[2] M. Bousquet-Mélou, A method for the enumeration of various classes of column-convex polygons, *Discrete Math.* **154** (1996) 1–25.

[3] M. Bousquet-Mélou and M. Petkovšek, in preparation.

[4] R. Cori and J. Richard, Énumération des graphes planaires à l'aide des séries formelles en variables non commutatives, *Discrete Math.* **2** No.2 (1972) 115–162.

[5] P. Duchon, On the enumeration and generation of generalized Dyck words, *Discrete Math.*, this volume.

[6] A. Dvoretzky and Th. Motzkin, A problem of arrangements, *Duke Math. J.* **14** (1947) 305–313.

[7] S. Feretić and D. Svrtan, On the number of column-convex polyominoes with given perimeter and number of columns, in: *Proceedings of the* 5^{th} *Conference "Formal Power Series and Algebraic Combinatorics"*, Florence, June 1993, pp. 201–214.

[8] P. Flajolet, Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.* **49** (1987) 283–309.

[9] I. M. Gessel, A factorization for formal Laurent series and lattice path enumeration, *J. Combin. Theory. Ser. A* **28** (1980) 321–337.

[10] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, John Wiley and Sons, New York, 1966.

[11] M. G. Kendall, *A Course in the Geometry of n Dimensions*, Hafner Publ. Co., New York, 1961.

[12] D. E. Knuth, *The Art of Computer Programming, Vol. 1: Fundamental Algorithms.* Addison-Wesley, Reading Mass., 1968.

[13] J. Labelle, Langages de Dyck généralisés, *Ann. Sci. Math. Québec* **17** (1993) 53–64.

[14] J. Labelle and Y.-N. Yeh, Generalized Dyck paths, *Discrete Math.* **82** (1990) 1–6.

[15] L. Lipshitz, D-finite power series, *J. Algebra* **122** (1989) 353–373.

[16] J. H. Loxton and A. J. van der Poorten, A class of hypertranscendental functions, *Aequationes Math.* **16** (1977) 93–106.

[17] P. Luksch and M. Petkovšek, An explicit formula for $|FM(1+1+n)|$, *Order* **6** (1990) 319–324.

[18] P. A. MacMahon, *Combinatory analysis*, 2 vols., Cambridge University Press, Cambridge, 1915–1916 (Reprinted: Chelsea, New-York, 1960).

[19] M. Petkovšek, *Finding closed-form solutions of difference equations by symbolic methods*, Ph.D. Thesis, Carnegie Mellon University, Pittsburgh PA, 1991 (CMU-CS-91-103).

[20] M. Petkovšek, The irrational chess knight, in: *Proceedings of the* 10^{th} *Conference "Formal Power Series and Algebraic Combinatorics"*, pp. 513–522, Toronto, June 1998.

[21] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton NJ, 1970.

[22] R. P. Stanley, Differentiably finite power series, *European J. Combin.* **1** (1980) 175–188.

[23] R. Suarez, Difference equations and a principle of double induction, *Math. Magazine* **62** (1989) 334–339.