The fine structure of 321 avoiding permutations.

M. H. Albert

July 10, 2004

Abstract

Bivariate generating functions for various subsets of the class of permutations containing no descending sequence of length three or more are determined. The notion of absolute indecomposability of a permutation is introduced, and used in enumerating permutations which have a block structure avoiding 321, and whose blocks also have such structure (recursively). Generalizations of these results are discussed.

1 Introduction

The association of a permutation of $\{1, 2, \ldots, n\}$ with its graph provides a geometric viewpoint in which to consider pattern avoidance. Thus, for example the permutation avoids 132 if the points to the left of the highest point, all lie above the points to the right, and this condition is true recursively of the points to the left and to the right of the highest point. This condition is illustrated in Figure 1. Likewise any set of n points in the plane, no two of which lie on a horizontal or vertical line can be associated with a permutation of $\{1, 2, \ldots, n\}$. In that case we can represent the geometric information about 132 avoidance as a picture, with a point for the maximum, and two bounding rectangles to its left and right, the former lying above the latter, together with an implicit understanding that the structure within the rectangles is to be similar. These associations provide a geometrical context in which to consider pattern avoidance, and are a common tool in understanding the class of permutations that avoid one or more patterns. We explore some ramifications of this viewpoint in the very simple situation of 321 avoiding permutations (which we denote by $A(321)$).

The main purpose of this paper then, is to show how the geometric context provides a simple method to obtain more detailed enumeration results about 321 avoiding permutations than have hitherto been available. Moreover, these results are obtained uniformly in some sense. The underlying technique consists of identifying a suitable geometric configuration which must be attained or avoided, and then using the structural constraints which that implies in order

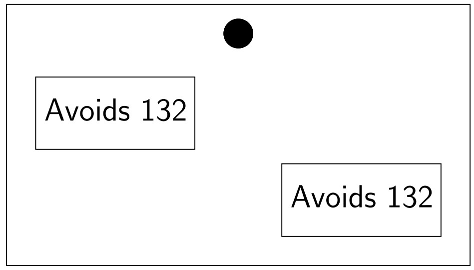

Figure 1: Schematic representation of a 132 avoiding permutation.

to compute the generating function, often multivariate, of the associated collection of permutations. This technique has applications beyond the scope of 321 avoidance, a few of which we consider in the final sections of the paper.

Thought of as a set of points, a permutation avoids 321 if it does not contain three points, every pair of which determines a line segment of negative slope. Of course it is also the case that any 321 avoiding permutation is the merge of two increasing sequences. It is easy to see that one of these sequences can be taken as the sequence of left to right maxima, that is, those elements which dominate all of their predecessors. These elements now determine a sequence of vertical and horizontal ranges in which the remaining elements of the permutation must lie. The situation is illustrated in Figure 1. Subject to having a fixed number of left to right maxima, the possible 321 avoiding permutations that can be formed are in one to one correspondence with assignments of non-negative integers to the cells of that diagram, so that no two cells whose centres are connected by a segment of negative slope are assigned positive labels. We refer to this diagram, with the cells containing positive labels simply marked (but the labels themselves suppressed) as the *skeleton* of the permutation.

As is well known, the total number of 132 avoiding permutations of length n and the total number of 321 avoiding permutations of length n are the same, both being equal to the nth Catalan number. However, the schematic representation of the 132 avoiding permutations makes the correspondence between them and plane binary trees clear and hence also the equation satisfied by the generating function of the class, while the corresponding diagram for the 321 avoiding permutations does not. In some sense $A(321)$ is a class which exhibits more subtle structure than $A(132)$ does.

Alternatively we might replace each cell in Figure 1 with its central point, producing a triangular subset of the integer grid. Given a 321 avoiding permutation, the points corresponding to cells with positive labels determine a path in this grid, lying on or below a diagonal line and having steps which are of the form (a, b) where a and b are non-negative integers, not both 0. This path is also shown in Figure 1. Such paths are enumerated in [15] and [14], the former of which also provides a bijective interpretation of the relationship between their

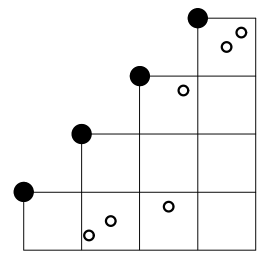

Figure 2: A 321-avoiding permutation with four left to right maxima. The given occupancies of the cells define the permutation **4** **5** 1 2 **7** 3 6 **10** 8 9. Note that the relative horizontal position of the occupants of the two vertically aligned cells is determined by the fact that the resulting permutation is not to contain 321.

enumeration and that of standard Delannoy and Schröder paths.

Our objective is to find the enumeration via generating functions of various subsets of the 321-avoiding permutations. Generally we will aim to obtain a multivariate generating function (in fact, at most bivariate) say in x and y where the coefficient of $x^n y^k$ might be the number of 321-avoiding permutations of a certain type having n left to right maxima and k other elements. Then simple substitutions will allow us to compute either the associated univariate generating functions, or permutations of size n having k left to right maxima, or other similar variations. Whenever possible, the name of a generating function will reflect the name of the class that it enumerates, so that for example the univariate generating function for the class $A(321)$ will be $A(t)$, while a bivariate form might be $A(x, y)$.

To carry out the enumerations of various structurally defined subsets of the collection of all 321 avoiding permutations we will make use of the structural relationships that hold between the subsets and the class as a whole, and the corresponding algebraic relationships which hold for the generating functions. That is, our methodology is firmly in the school represented by [10], or [8].

In the next section we rederive equations describing the basic generating functions for 321 avoiding permutations by defining a context-free language whose elements are in one to one correspondence with 321 avoiding permutations. This method is used because it is no longer than extending the results of [15] to the bivariate case, and also because it represents a technique of wider applicability in the field of pattern class enumeration.

The following section then applies the results to the problem of enumerating the subsets of 321 avoiding permutations consisting of: plus irreducible, minus irreducible, plus indecomposable, and absolutely irreducible permutations. These terms, and their significance for enumeration questions are defined below. In the univariate case, only the last of these is definitely new, although we have not found detailed expositions of the others in the literature.

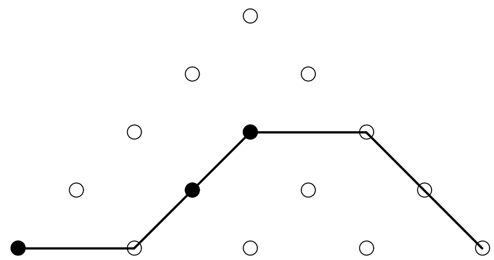

Figure 3: An illustration of the correspondence between occupied cells in the triangular grid and lattice paths. Points in the subset are marked as solid circles, and the corresponding path is illustrated. The associated word is mhumumhdd.

In the penultimate section we make use of the enumeration of the absolutely irreducible 321 avoiding permutations to enumerate another class, built from 321 avoiding permutations by a recursive construction based upon the wreath product introduced in [3].

In the final section we try to foreshadow future applications of the methods illustrated here, and mention some connections with other work in the area of pattern classes.

2 Enumeration of marked paths

In Figure 2 we rotate the triangular grid from Figure 1 clockwise by 45°, and then reflect it through the x-axis. The paths we then obtain begin from $(0, 0)$, and travel through lattice points (the sum of whose coordinates is even) along segments of slope lying between -1 and 1 inclusive. We will introduce a grammar which describes such paths. The purpose of transforming the original diagram is purely psychological. The grammar which we describe is the language accepted by a certain deterministic pushdown automaton, and the lattice path then models the number of elements held in the stack of this automaton as the word is processed. Given our overlying motivation of considering pattern classes this also provides a nice link to the generation of 312-avoiding permutations using a stack.

Allowing path segments of arbitrary rational slope in $[-1, 1]$ would obviously be a problem in a finite language. We avoid this problem by adding marks to the end of each path segment, and then replacing each segment by a horizontal segment followed by one at an inclination of $\pm 45°$. This is illustrated in Figure 2.

We now describe a context free grammar which generates a language that describes all, and only, the lattice paths which correspond to marked subsets of the triangular grid of size n for some n. Constructing this grammar is relatively straightforward. We use four terminal symbols, three of which stand for unit segments in the path, and one of which represents a vertex in the set: h for **h**orizontal segments, u for **u**pward segments, d for **d**ownward segments, and m for **m**arking a vertex of the original path.

Informally words in the language are described as follows:

- Any change of direction after a d or a u requires an intervening m.
- h's may only occur in blocks immediately following an m or at the beginning of a word.
- In any initial segment there must be at least as many u's as d's, but in the whole word the total number of each is the same.
- m's can occur anywhere.

This description is easily formatted into a grammar, modelled on the standard grammar corresponding to Dyck paths (which do not allow horizontal steps, and do not require marks). Each non-terminal symbol in the grammar represents an excursion, that is a path starting and ending at the same level and not passing below that level. These excursions are sometimes restricted by the immediately preceding symbol.

$$\begin{aligned} S &\rightarrow \epsilon \mid \mathsf{h}S \mid \mathsf{u}U\mathsf{d}D \mid \mathsf{m}M \\ U &\rightarrow \mathsf{m}M \mid \mathsf{u}U\mathsf{d}D \\ D &\rightarrow \epsilon \mid \mathsf{m}M \\ M &\rightarrow \epsilon \mid \mathsf{h}S \mid \mathsf{u}U\mathsf{d}D \end{aligned}$$

As in [6] each non-terminal symbol of the grammar is associated with a generating function (denoted by the same symbol) in variables h, u, d, m. This generating function is obtained by taking the sum of the monomials corresponding to words represented by that non-terminal.

The grammar above is clearly unambiguous since in each rule the initial symbols of differing productions differ from one another. So, it is a simple matter to obtain a system of equations satisfied by the generating functions of the non-terminals, namely:

$$\begin{aligned} S &= 1 + hS + udUD + mM \\ U &= mM + udUD \\ D &= 1 + mM \\ M &= 1 + hS + udUD. \end{aligned}$$

Using a symbolic algebra package or, in a pinch, by hand, this system can be solved. We are interested principally in the function S describing words of the language, and this is described as $S = (1+m)S_1$ where:

$$\begin{array}{rr} \left(-udm^2 - udm + udmh + udm^2h\right) S_1^2 & + \\ (1 - h - mh - ud + udh + udmh)\, S_1 - 1 + ud & = 0. \end{array}$$

Given a word in the language, the number of left-to-right maxima in the 321 avoiding permutation of which it is the skeleton is equal to one more than the sum of the number of h's and the number of u's. Also, the size of the set that it encodes is equal to the number of m's. So, we can reduce to the generating function $S(x, y)$ where the coefficient of $x^n y^k$ is the number of skeletons with n left to right maxima, and k internal marked cells through the following substitutions:

$$h \to x,\ u \to x,\ d \to 1,\ m \to y,$$

followed by multiplication by x (and addition of 1 for the empty graph). This yields:

$$\begin{array}{rcl} S(x, y) & = & 1 + x(1 + y)S_1(x, y) \\ 0 & = & (xy + xy^2)S_1^2 + (xy + x - 1)S_1 + 1. \end{array}$$

The first of these equation can be solved for S_1, with the result being substituted in the second. After some further simplification this yields:

$$xS^2 + (xy + x - 2y - 1)S + 1 + y = 0. \quad (1)$$

By substituting $y = 1$ we will obtain the total number of allowed markings. So, defining $S(x) = S(x, 1)$:

$$S(x)^2 + (2x - 3)S(x) + 2 = 0.$$

The discriminant of the latter equation is $\sqrt{4x^2 - 12x + 1}$ illustrating a connection between these numbers and the Schröder numbers (sequence A001003 of [13]). In fact:

$$S(x) = 1 + 2x + \sum_{n=1}^{\infty} 2^{n+1} s_n x^n$$

where s_n is the nth Schröder number.

This sequence of coefficients also arises as the number of *non-crossing graphs*, that is, graphs with n vertices arranged as the vertices of a convex polygon, with straight edges connecting these vertices subject to the condition that no two edges should intersect at an interior point. This result is due to [7], with a more modern derivation, as well as other related results, given in [9].

If we consider the $\binom{n+1}{2}$ cells of the original grid as the vertices of a graph, T_n, two vertices being adjacent if they are connected by a line of negative slope, then the coefficient of x^n in $S(x)$ counts the independent subsets of T_n. The number of non-crossing graphs is also the number of independent sets in a graph.

Namely, take as vertices of the graph the possible edges two such vertices being adjacent if the segments which they represent meet internally. In this graph NC_n, a non-crossing graph corresponds to an independent set.

So, the generating function for independent subsets of the sequence of graphs T_n and NC_{n+1} are the same. In fact, inspection of the results in [9] together with a little algebra shows that this is also true of the bivariate generating functions which mark the sizes of the independent subsets. That is:

Proposition 1 *For every n and every k, T_n and NC_{n+1} have exactly the same number of independent subsets of size k.*

However, it is easy to see that for $n \geq 4$, T_n and NC_{n+1} are not isomorphic. For, T_n has exactly four isolated vertices, while NC_{n+1} has $n+1$ isolated vertices. Detailed expressions for the coefficients in $S(x)$ and $S(x, y)$ can be found in [9] (Theorem 2, part (ii)) as well as discussions of their asymptotic expansions.

3 Consequences for 321 avoiding permutations

Before turning to the enumeration of various subsets of the 321 avoiding permutations we begin with some remarks about the full class (whose enumeration is, of course, already well understood beginning apparently from [11]).

In our original setting, the marked cells arose by considering a 321 avoiding permutation having n left-to-right maxima. We argued that any such permutation corresponded to a labelling of marked cells with positive integers representing the number of elements of a permutation contained in a particular cell. Let $A(x, y)$ be the generating function for 321 avoiding permutations where the exponent of x denotes the number of left to right maxima, and that of y the number of remaining elements. Since we obtain a 321 avoiding permutation from its skeleton by replacing a single cell, marked by a y in $S(x, y)$, by a positive integer, marked therefore by y^n for some $n > 0$, we obtain:

$$A(x, y) = S(x, y + y^2 + \cdots) = S\left(x, \frac{y}{1-y}\right)$$

We can also make this substitution in the equation that S satisfies and then simplify to obtain:

$$yA^2 + (x - y - 1)A + 1 = 0. \tag{2}$$

On the other hand, it is perhaps more natural to count permutations of a common size. So, using A_{am} to denote the generating function where the coefficient of $x^n y^k$ is the number of 321 avoiding permutations of length n having k left to right maxima, we obtain:

$$A_{am}(x, y) = S\left(xy, \frac{x}{1-x}\right).$$

By algebraic manipulation this function also satisfies a quadratic equation with coefficients polynomial in x and y namely:

$$xA_{am}^2 + (xy - x - 1)A_{am} + 1 = 0 \tag{3}$$

A further reduction in complexity occurs when we substitute $y = 1$ in A_{am} (or $y = x$ in (2)) giving:

$$xA(x)^2 - A(x) + 1 = 0$$

thus confirming, in a rather roundabout way, that the total number of 321 avoiding permutations of length n is enumerated by the Catalan numbers.

The coefficient of $x^n y^k$ in $A_{am}(x, y)$, which is non-zero only for $1 \leq k \leq n$ is a *Narayana number*,

$$[x^n y^k]A = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}.$$

These numbers also arise in [15], but not as a direct translation of this result since we are no longer in the context of path counting. They also arise in a number of other contexts including the enumeration of k-way trees ([2]) and as the number of non-crossing partitions of n ([9]).

We now turn to the enumeration of various subsets of $A(321)$. First let us define those classes and the symbols used to specify them:

Definition 2 *Let π be a permutation (in $A(321)$). Then:*

(A_{+irr}) *π is* plus irreducible *if it does not contain a subword of the form $i\,(i+1)$,*

(A_{-irr}) *π is* minus irreducible *if it does not contain a subword of the form $i\,(i-1)$,*

(A_{+ind}) *π is* plus indecomposable *if it does not have a proper initial segment whose values form an initial segment of $[1, n]$,*

(A_{-ind}) *π is* minus indecomposable *if it does not have a proper final segment whose values form an initial segment of $[1, n]$,*

(A_{irr}) *π is* absolutely irreducible *if it does not have a proper subword of length greater than 1 whose values form an interval in $[1, n]$.*

The irreducible or indecomposable elements of a collection of permutations can (under suitable closure properties) be thought of as components in the construction of the other elements of that class. Again, granted certain closure and uniqueness assumptions, this can allow enumeration of the entire set based on an enumeration of one of the collections of components, or vice versa. This particular exposition of a general combinatorial theme is explored in [3]. We note that the results in that paper could be used to derive the univariate generating function for the plus irreducibles and plus indecomposables in $A(321)$

(results which we will rederive here as a result of obtaining the bivariate form). Furthermore, the only minus decomposable permutations that avoid 321 are of the form:

$$(k+1)\,(k+2)\cdots n\,1\,2\cdots k$$

so we will not concern ourselves with that case.

The condition of absolute irreducibility is a new one, and we will see its application in the next section. The definition is not so unnatural as it might appear to be at first sight. In terms of the graph of a permutation it says that if some proper, non-singleton, part of the permutation is bounded by a rectangle, then there must be at least one element of the permutation outside of the rectangle but in either the vertical strip or the horizontal strip determined by it.

Enumeration results in this section generally take equation (1) as their starting point. Recall that this provides the generating function $S(x, y)$ for skeletons of 321 avoiding permutations, with the exponent of x marking the number of left to right maximals, and that of y the number of occupied cells. So, all the generating functions we compute will be in the form where the coefficient of $x^n y^k$ marks the number of permutations of that type having n left to right maxima and k other elements. As usual, a simple change of variable, replacing x by xy and y by x would produce the function enumerating by total number of elements, and number of left to right maxima.

If π is a plus irreducible member of $A(321)$ then no cell can be occupied by more than one element. Among the diagrams that meet this criteria, the plus reducible elements contain sequences of more than one left to right maximum such that the vertical and horizontal bands which they determine are otherwise empty. Suppose then that we knew the generating function $A_{+irr}(x, y)$ for the plus irreducible members of the class. The preceding sentences imply that we would obtain the generating function $S(x, y)$ by replacing x in $A_{+irr}(x, y)$ by $x/(1-x)$. So, since the inverse of sending x to $x/(1-x)$ is to send it to $x/(1+x)$:

$$A_{+irr}(x, y) = S(x/(1 + x), y).$$

Substitution and simplification in equation 1 then yields:

$$x(y + 1)A_{+irr}^2 - (xy + 2y + 1)A_{+irr} + (x + 1)(y + 1) = 0. \tag{4}$$

The corresponding univariate form is:

$$x(x + 1)A_{+irr}^2 - (x + 1)^2 A_{+irr} + (x + 1)^2 = 0.$$

An element π of $A(321)$ can only be minus reducible if some left to right maximum k is followed immediately by $k - 1$. So

$$\pi = \alpha k\,(k - 1)\,\beta,$$

for some $\alpha, \beta \in A(321)$ (with β of course having all its values increased by k). We can make this decomposition unique by requiring k, $k-1$ to be the first pair of elements witnessing minus reducibility. Then α is minus irreducible, while β could be any 321-avoiding permutation. Thus we obtain:

$$A(x, y) = A_{-irr}(x, y) + A_{-irr}(x, y)(xy)A(x, y).$$

Or, solving for $A_{-irr}(x, y)$:

$$A_{-irr}(x, y) = \frac{A(x, y)}{1 + xyA(x, y)}. \tag{5}$$

The bivariate algebraic equation for A_{-irr} is not very pretty, but the univariate form is more presentable:

$$(x^4 + x^2 + x)A_{-irr}^2 + (1 - 2x^2)A_{-irr} + 1 = 0.$$

Enumerating plus indecomposables is easier and standard. Every element of $A(321)$ is either of length 0 or of the form $\alpha_1\alpha_2 \cdots \alpha_c$ where each α_i is a plus indecomposable, shifted upwards by the sum of the lengths of the preceding α's. Since this decomposition is unique, then using A_{+ind} to enumerate the non empty plus indecomposables, we obtain:

$$A = \frac{1}{1 - A_{+ind}},$$

which can then be readily solved for A_{+ind}.

Finally we come to absolute irreducibility. Since the absolutely irreducibles form a subset of the collection of plus indecomposables, and of the plus irreducibles, we begin with the form of the skeleton function which is like that for plus indecomposables. This already reduces us to permutations that are plus indecomposable, and plus irreducible in their non left to right maxima.

$$S_{+ind}(x, y) = \frac{S(x, y)}{1 + S(x, y)}.$$

Which, by now standard manipulations, satisfies:

$$(1 + y)S_{+ind}^2 - (1 + x + xy)S_{+ind} + xy + x = 0.$$

Consider which non empty rectangles in the diagram associated to an element of S_{+ind} might not contain other elements inside the vertical and horizontal strip which they define. In order for this to hold, the top edge of the rectangle cannot cross a vertical line in the triangular grid of cells, nor can the left edge cross such a horizontal line. So, the upper right and lower left corners lie outside of the grid. Such a rectangle is illustrated in Figure 3. The vertical area above the rectangle is automatically empty as is the horizontal area to the left. So problems can occur only when we have a non-empty sequence of left to right maxima such that there are no marked cells in the horizontal or vertical strip which they define.

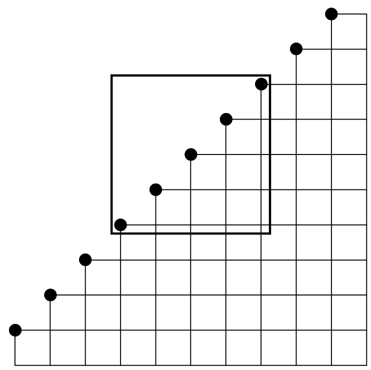

Figure 4: A potential rectangle for the violation of indecomposability.

If we knew the function $A_{irr}(x, y)$ how could we compute $S_{+ind}(x, y)$? An element of the latter class could be obtained beginning from an element of $A_{irr}(x, y)$ by inflating some of the left to right maxima into a sequence of such maxima, adding no additional elements in the horizontal or vertical strips which they determine. If we imagine in Figure 3 that the illustrated rectangle (and subrectangles of it) are the only ones which cause a violation of absolute indecomposability, then that permutation has been constructed by inflating the left to right maximum just to the left of the rectangle into six such maxima. As we've already insisted on plus indecomposability, a rectangle whose leftmost boundary is to the left of the first maximal cannot be problematic, and so there always is an available maximal to inflate.

There is just a single exception. The permutation 1 is absolutely indecomposable, but when we inflate it we do not obtain plus indecomposable permutations.

Thus, beginning from $A_{irr}(x, y) - x$, we should replace x by $x/(1-x)$ in order to obtain $S_{+ind}(x, y)$. Inverting this replacement we get:

$$A_{irr}(x, y) = S_{+ind}\left(\frac{x}{1+x}, y\right) - \frac{x}{1+x}.$$

Carrying out these substitutions and manipulations on the equations satisfied by the generating function yields:

$$(1+x)(1+y)A_{irr}^2 + (xy-1)A_{irr} + xy = 0. \qquad (6)$$

Or in univariate form:

$$(x+1)^2 A_{irr}^2 + (x^2-1)A_{irr} + x^2 = 0.$$

In fact, this does not quite get us *all* the irreducibles as it omits 1, 12, and the empty permutation. Adjusting the equations to include this one comes at considerable cost to their appearance, so we prefer to leave the equation as it stands, adding the necessary $1 + x + x^2$ to the generating function *post facto*. Table 3 summarizes the sizes of these subsets of $A(321)$.

n	0	1	2	3	4	5	6	7	8	9	10
all	1	1	2	5	14	42	132	429	1430	4862	16796
plus irr.	1	1	1	2	4	9	21	51	127	323	835
minus irr.	1	1	1	3	10	31	98	321	1078	3686	12789
abs. irr.	1	1	2	0	2	2	7	14	37	90	233

Table 1: Sizes of $A(321)$ and some of its subsets

.

4 A "fractal" class

As an application of the final results of the preceding section we will show how the knowledge of the generating function for A_{irr} can be used to compute that of a much more complicated class. At the risk of further abusing a term which has suffered much abuse already we would like to introduce the class $F(321)$ of *fractal* 321-*avoiders*. These are permutations which, from a distance, appear to avoid 321 but which on closer inspection are made up of blocks, arranged in a 321-avoiding pattern where each block appears to avoid 321 but perhaps on closer inspection is in fact made up out of blocks ...

That is, $\pi \in F(321)$ if either, $\pi \in A(321)$, or $\pi = \alpha_1 \alpha_2 \cdots \alpha_c$ where

- the values occurring in each α_i form an interval,
- the permutation represented by α_i is in $F(321)$, and
- the relative ordering of the α_i, interpreted as a permutation of length c is in $A(321)$.

Many well-known permutation classes can be defined as fractal classes in this way, or occasionally as natural subclasses of such fractal classes. For example, the class of separable permutations is precisely the fractal class generated from the finite base class $\{12, 21\}$.

There is a complementary bottom up description of $F(321)$. Namely, this class is the closure of the class consisting just of 1 under the operation of replacing an element of a permutation by a 321-avoiding block. For example:

$$1 \to 2413 \to 423615$$

where we initially replace 1 by 2413 and then replace the element 2 by the permutation 312 while retaining its relative order within the entire permutation. Geometrically, we begin with the graph of 2413 and then expand the vertex representing 2 into a copy of the graph of 312. Such replacements could just as easily be applied to each element of a permutation and, in some sense, they already have been, only 1 has been replaced by 1 in three instances. Thus the two descriptions are equivalent – the permutation 423615 consists of blocks

(423)(6)(1)(5) whose relative order is 321-avoiding, and where each block is in $F(321)$ (in this instance, in fact in $A(321)$).

We could also define $F(321)$ algebraically using the wreath product operator of [3] as the smallest non-empty class X satisfying the equation $X = X \wr A$ where $A = A(321)$. This corresponds to the bottom up description, while the top down one would suggest $X = A \wr X$. Consider the first equational description of X. Since X contains A we also get that X contains $A \wr A$. Then also X contains $(A \wr A) \wr A$ and so on. Letting $A^n = A^{n-1} \wr A$ for $n > 1$, and $A^1 = A$ we obtain

$$X \supseteq \cup_{n=0}^{\infty} A^n.$$

On the other hand, the right hand side is contained in its wreath product with A, and so by the definition of X:

$$X = \cup_{n=0}^{\infty} A^n.$$

The second equational definition can be manipulated in the same way and in fact, as the wreath product is associative, leads to the same equation, thus confirming that the two approaches are indeed equivalent.

Such an algebraic representation suggests that we ought to be able to transfer our knowledge of generating functions for A to similar knowledge about F. There is though, a small complication. This arises from the fact that the choice of blocks to witness the fact that a permutation belongs to $F(321)$ is not uniquely defined. We need to obtain uniqueness of some sort if we hope to carry out the enumeration, and the following general result helps to provide that.

Definition 3 *Let θ be a permutation of length k and let π be a permutation of length n. Then π is θ-*decomposable *if $\pi = \alpha_1\alpha_2 \cdots \alpha_k$ for some non-empty subwords α_i such that the set of values occurring in each of the α_i forms an interval and the relative ordering of these values agrees with the relative ordering of the corresponding elements of θ. The factorization $\pi = \alpha_1\alpha_2 \cdots \alpha_k$ is called a θ-*decomposition *of π.*

With this new definition, we see that a permutation of length 3 or more is plus decomposable if and only if it is 12-decomposable, while a permutation π is absolutely irreducible if and only if it is not θ-decomposable for any $\theta \neq \pi$.

Proposition 4 *Let π be an arbitrary permutation. Then there is a unique absolutely irreducible permutation θ such that π is θ-decomposable. Moreover, if $\theta \neq 12$ and $\theta \neq 21$ then the θ-decomposition of π is also unique.*

For example, for 423615 this decomposition is (423)(6)(1)(5) with relative ordering 2413, while for 724513986 it is (7)(24513)(98)(6) with relative ordering 3142. On the other hand 123 which is 12-decomposable admits two such decompositions.

Proof: Let π be given, say $\pi = p_1 p_2 \cdots p_n$. To each p_i associate a maximal proper subword α_i of π such that the values occurring in α_i form an interval (of course, α_i might well be a singleton).

Suppose that there are $i < j$ such that α_i and α_j overlap properly but are not equal. Then the elements of π belonging to either α_i or α_j form a subword α whose values are the union of two overlapping intervals, hence an interval. By the maximality of either α_i or α_j it must be the case that $\alpha = \pi$. Then α_i and α_j with the elements common to α_i deleted form either a 12 or a 21 decomposition of π. These cases are clearly mutually exclusive.

Henceforth suppose that π is neither 12-decomposable nor 21-decomposable. Then the α_i form a partition of π (i.e. any two are either equal or disjoint). The relative ordering of the α_i must be some absolutely irreducible permutation θ, for otherwise we could pool some proper subset of the α's to form a coarser partition, contradicting the choice of each α_i as a maximal proper subword of π whose values form an interval. Now reindex the distinct α_i and write $\pi = \alpha_1 \alpha_2 \cdots \alpha_k$.

Thus we have established the existence of a decomposition of the type claimed. To establish uniqueness, suppose that another decomposition of the same kind, say $\pi = \beta_1 \beta_2 \cdots \beta_m$ were given. We include here the assumption that the relative ordering of β_1 through β_m forms an absolutely irreducible permutation. If $\beta_1 \neq \alpha_1$ then β_1 is a subword of α_1 by the maximality of α_1. Now take the least j such that $\beta_1 \beta_2 \cdots \beta_j$ contains α_1. Then in fact we must have $\beta_1 \beta_2 \cdots \beta_j = \alpha_1$ for otherwise the values in α_1 and β_j form overlapping intervals, and so the values in $\beta_1 \beta_2 \cdots \beta_j$ form an interval, contradicting the maximality of α_1 ($\beta_1 \beta_2 \cdots \beta_j \neq \pi$ since π is neither 12 nor 21-decomposable). However, $\beta_1 \beta_2 \cdots \beta_j = \alpha_1$ contradicts the absolute irreducibility of the relative ordering of the β's. So $\alpha_1 = \beta_1$. But now the same argument implies that $\alpha_2 = \beta_2$ and, inductively that in fact $m = n$ and $\alpha_i = \beta_i$ for all i. ■

We now return to the analysis of $F(321)$. Let $\pi \in F(321)$ be given. Suppose that it is neither 12-decomposable nor 21-decomposable. By the proposition above, $\pi = \alpha_1 \alpha_2 \cdots \alpha_k$ for some subwords α_i whose values form intervals, and whose relative ordering forms an absolutely irreducible permutation. Since π has some decomposition into subwords whose relative ordering avoids 321, and since the proof of the proposition above shows that the α's form the coarsest possible proper partition of π into subwords whose values form intervals, it must be the case that the relative ordering of the α_i avoids 321. Of course, we also have that each α_i belongs to $F(321)$. Conversely, given α_i in $F(321)$, shifted to have relative order equal to some absolutely irreducible element θ of $A(321)$, then, by the very definition of $F(321)$, the permutation $\alpha_1 \alpha_2 \cdots \alpha_k$ belongs to $F(321)$.

Let $F(x)$ be the generating function for $F(321)$, taken to have constant term 0, and $A_i(t)$ be the univariate generating function for the absolutely indecomposable members of $A(321)$ of length greater than or equal to 3. From the above we see that $A_i(F(x))$ is the generating function for the elements of F of length at least 3 which are neither 12-decomposable nor 21-decomposable. Let F_+ denote the generating function for the 12-indecomposable elements of F, and F_- that of the 21-indecomposable elements of F, again taken with constant term 0. Then F_+F enumerates the 12-decomposable members of F while F_-F enumerates the 21-decomposables. Further relations arise from the observation that a 12-indecomposable is either 21-decomposable or both 12- and 21-indecomposable, and similarly for minus indecomposables. We thereby obtain the system of equations:

$$\begin{aligned} F &= x + F_+F + F_-F + A_{irr}(F) \\ F_+ &= x + F_-F + A_{irr}(F) \\ F_- &= x + F_+F + A_{irr}(F). \end{aligned}$$

Solving this system for F gives:

$$F^2 + (A_{irr}(F) - 1 + x)F + A_{irr}(F) + x = 0. \tag{7}$$

We can use the work of the previous section to obtain a radical expansion of A_{irr}:

$$A_{ind}(x) = \frac{1 - x - \sqrt{-3\,x^2 - 2\,x + 1}}{2(x+1)} - x^2.$$

Then substitution in (7) and elimination of radicals gives:

$$\begin{aligned} F^6 + (-2\,x + 3)\,F^4 + (-2\,x - 1)\,F^3 + \\ \left(-3\,x + 3 + x^2\right)F^2 + \left(2\,x^2 - 1 - 2\,x\right)F + x + x^2 &= 0. \end{aligned} \tag{8}$$

The first few terms of the associated power series are:

$$x + 2\,x^2 + 6\,x^3 + 24\,x^4 + 116\,x^5 + 625\,x^6 + 3580\,x^7 + \\ 21297\,x^8 + 130084\,x^9 + 810737\,x^{10} + O\left(x^{11}\right)$$

and the exponential constant governing the growth rate, is the reciprocal of the radius of convergence of this series. This radius is the least positive root of the discriminant of (8), which is an irreducible polynomial of degree 7. The value of the exponential constant is approximately 7.346751, compared to 4 for the underlying class $A(321)$.

5 Summary and Conclusions

We began this paper with an explicitly constructed grammar to describe the skeletons of 321 avoiding permutations. In general there is a close connection between combinatorial classes with algebraic generating functions and unambiguous context free languages. This connection can either be used, as here, to provide an explicit enumeration of a class, or to provide a "soft" proof that the generating function of a class is in fact algebraic. The former approach has become much more attractive with the ready availability of symbolic algebra packages since the algebraic manipulations necessary to solve the equations arising from the grammar are undeniably tedious. The latter approach has been used in [1] to provide algebraicity results for a family of pattern classes. It can also be used in the context of generating functions for generating trees, thereby generalizing a number of the theorems in [4] about the existence of algebraic generating functions. It must be noted though that the results of that paper and similar results in [5] provide much more explicit detail concerning the generating functions that they produce.

One of the striking features of the equations for the various irreducible and indecomposable subsets of $A(321)$ is their simplicity. In some sense then the enumerative coincidences that we observed are not so startling, since there is a relatively limited supply of simple quadratic equations. Because of the simple form of these equations, one could apply Lagrange inversion to obtain explicit formulae for many of the coefficients, as is done for example in [9], or indeed carry out detailed asymptotic analyses of these coefficients.

We restricted ourselves to bivariate generating functions but the reader should note that the techniques employed can be naturally applied to produce other statistics of these permutations. For example, it would be a simple matter to produce, if one wished, a generating function $A(x, y, z, w)$ where x marked total size, y marked left to right maxima, z marked the number of occurences of $i(i + 1)$ among the left to right maxima, and w that number among the remaining elements.

The class $A(321)$ is the simplest pattern class, in terms of the patterns which it avoids, that contains infinitely many absolutely irreducibles. The techniques used in the preceding section to solve (in the sense of enumeration) the wreath fixed point equation:

$$X = A \wr X$$

apply, owing to proposition 4, completely generally to any base class A in which the absolutely irreducibles can be enumerated. In particular for a positive integer n let D_n be the class of permutations which "fractally have $\leq n$ elements". That is, they are comprised of at most n blocks, each of which is comprised of at most n blocks, each of which ... Then D_n is the solution of the fixed point equation:

$$X = F_n \wr X$$

where F_n is the class of permutations of size $\leq n$. Since F_n is finite equation (7) is simply a polynomial and we obtain:

Corollary 5 *Each of the classes D_n has an algebraic generating function.*

There is much further information to be gleaned from the representation of a class as a subclass of D_n when this is possible, and we hope to explore these matters in a future paper.

One aspect of $F(321)$ that has been notably omitted is a description in terms of minimal forbidden patterns. It appears that this set may be finite consisting of:

$$42513, 35142, 41352, 362514, 531642$$

but all that can be said with certainty at this point is that no further minimal forbidden patterns exist of length 12 or less.

6 Acknowledgements

Many of the connections between these problems and known results would have been missed without the electronic version of the encyclopedia of integer sequences ([12]). Mike Atkinson listened patiently to many preliminary, and incorrect, expositions, and also provided a key step in the proof of Proposition 4. *Maple* did most of the hard work.

References

[1] M.H. Albert and M.D. Atkinson. Sorting with a forklift. *Otago University CS Tech. Report*, 2002-06:25pp, 2002.

[2] M. D. Atkinson and Louise Walker. Enumerating k-way trees. *Inform. Process. Lett.*, 48(2):73–75, 1993.

[3] M.D. Atkinson and T. Stitt. Restricted permutations and the wreath product. *Discrete Math.*, 259:19–36, 2002.

[4] Cyril Banderier, Mireille Bousquet-Mélou, Alain Denise, Philippe Flajolet, Danièle Gardy, and Dominique Gouyou-Beauchamps. Generating functions for generating trees. *Discrete Math.*, 246(1-3):29–55, 2002.

[5] Cyril Banderier and Philippe Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoret. Comput. Sci.*, 281(1-2):37–80, 2002.

[6] N. Chomsky and M. P. Schützenberger. The algebraic theory of context-free languages. In *Computer programming and formal systems*, pages 118–161. North-Holland, Amsterdam, 1963.

[7] C. Domb and A. J. Barrett. Enumeration of ladder graphs. *Discrete Math.*, 9:341–358, 1974.

[8] P. Flajolet and R. Sedgewick. *Analytic Combinatorics—Symbolic Combinatorics*. Preprint, http://algo.inria.fr/flajolet/Publications/books.html, 2002.

[9] Philippe Flajolet and Marc Noy. Analytic combinatorics of non-crossing configurations. *Discrete Math.*, 204(1-3):203–229, 1999.

[10] I. P. Goulden and D. M. Jackson. *Combinatorial enumeration.* A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1983.

[11] J. M. Hammersley. A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, pages 345–394, Berkeley, Calif., 1972. Univ. California Press.

[12] N. J. A. Sloane, editor (2002), The On-Line Encyclopedia of Integer Sequences, published electronically at `http://www.research.att.com/ njas/sequences/`.

[13] N. J. A. Sloane and Simon Plouffe. *The encyclopedia of integer sequences.* Academic Press Inc., San Diego, CA, 1995.

[14] Robert A. Sulanke. A symmetric variation of a distribution of Kreweras and Poupard. *J. Statist. Plann. Inference*, 34(2):291–303, 1993.

[15] Robert A. Sulanke. Counting lattice paths by Narayana polynomials. *Electron. J. Combin.*, 7(1):Research Paper 40, 9 pp. (electronic), 2000.