

Pattern avoidance classes and subpermutations

M. D. Atkinson

Department of Computer Science
University of Otago, New Zealand
`mike@cs.otago.ac.nz`

M. M. Murphy and N. Ruškuc

School of Mathematics and Statistics
University of St Andrews, Scotland, KY16 9SS
`nik@mcs.st-and.ac.uk` and `max@mcs.st-and.ac.uk`

Submitted: Oct 6, 2005; Accepted: Nov 10, 2005; Published: Nov 15, 2005

Mathematics Subject Classifications: 05A15, 05A16

Abstract

Pattern avoidance classes of permutations that cannot be expressed as unions of proper subclasses can be described as the set of subpermutations of a single bijection. In the case that this bijection is a permutation of the natural numbers a structure theorem is given. The structure theorem shows that the class is almost closed under direct sums or has a rational generating function.

Keywords: Restricted permutations, pattern avoidance, subpermutations.

1 Introduction

Classes of permutations defined by their avoiding a given set of permutation patterns have been intensively studied within the last decade. Quite often the issue has been to determine the number of permutations of each length in the class. In order to do this it is necessary to derive structural properties of the permutations in the class starting from the avoided set. However, there are very few general techniques for obtaining such structural information. This paper is a contribution towards a general structure theory. We begin from the point of view that pattern-avoidance classes can be expressed as unions of *atomic* classes (those that have no non-trivial expression as a union). We shall show that these atomic classes are precisely the classes that arise as the set of restrictions of some injection from one ordered set to another. In general the order types of these two sets provide some information about the atomic class. The major part of our paper is a

characterisation of such injections and classes in the simplest case: when the order types are those of the natural numbers.

In the remainder of this section we review the terminology of pattern avoidance classes. Most of this terminology is standard in the subject except for the notion of a *natural* class. We shall see a number of conditions on pattern avoidance classes that are equivalent to their being atomic and we shall exhibit examples of atomic and non-atomic classes. These conditions and examples motivate the notion of a natural class whose elementary properties we explore in Section 2. Section 3 contains our main result: a characterisation of natural classes, and Section 4 gives some further examples.

We need a small number of definitions concerned with permutations and sets of permutations. For our purposes a permutation is just an arrangement of the numbers $1, 2, \dots, n$ for some n , and we shall write these as lists of numbers (sometimes with separating commas to avoid notational confusion). We shall often need to consider arrangements of other sets of numbers and we shall refer to these as sequences; so, unless stated otherwise, a *sequence* will mean a list of *distinct* numbers.

Two finite sequences of the same length $\alpha = a_1a_2a_3 \dots$ and $\beta = b_1b_2b_3 \dots$ are said to be *order isomorphic* (denoted as $\alpha \cong \beta$) if for all i, j we have $a_i < a_j$ if and only if $b_i < b_j$. Any sequence defines a unique order isomorphic permutation; for example $7496 \cong 3142$.

A sequence α is said to be *involved in* a sequence β (denoted as $\alpha \preceq \beta$) if α is order isomorphic to a subsequence of β . Usually, involvement is defined between permutations; for example $1324 \preceq 146325$ because of the subsequence 1435 .

It is easily seen that the involvement relation is a partial order on the set of all (finite) permutations. We study it in terms of its order ideals which we call *closed sets*; see [8] for some similar definitions. A closed set X of permutations has the defining property that if $\alpha \in X$ and $\delta \preceq \alpha$ then $\delta \in X$.

Closed sets are most frequently specified by their *basis*: the set of permutations that are minimal subject to not lying in the closed set. Once the basis B is given the closed set is simply

$$\{\sigma \mid \beta \not\preceq \sigma \text{ for all } \beta \in B\}$$

and we shall denote it by $\mathcal{A}(B)$.

Closed sets arise in the context of limited capability sorting machines such as networks of stacks, queues and dequeues with a point of input and a similar output. Here the basis consists of minimal sequences that cannot be sorted into some desirable order. As sequences can be sorted if and only if they do not involve any basis elements, basis elements are frequently referred to as “forbidden patterns” and the set of permutations that can be sorted by such a mechanism is described as the set of all permutations that “avoid” that basis.

Much of the thrust of this paper is in specifying closed sets in a different way. Suppose that A and B are sets of real numbers and let π be an injection from A to B . Then every finite subset $\{c_1, c_2, \dots, c_n\}$ of A , where $c_1 < c_2 < \dots < c_n$ maps to a sequence

$\pi(c_1) \pi(c_2) \dots \pi(c_n)$ which is order isomorphic to some permutation. The set of permutations that arise in this way is easily seen to be closed and we denote it by $\text{Sub}(\pi : A \rightarrow B)$. In many cases the domain and range of π are evident from the context in which case we write simply $\text{Sub}(\pi)$. Also, since we may always replace B by the range of π we shall, from now on, assume that π is a bijection.

Example 1.1 Let $A = \{1 - 1/2^i, 2 - 1/2^i \mid i = 1, 2, \dots\}$ and $B = \{1, 2, \dots\}$. Let π be defined by:

$$\pi(x) = \begin{cases} 2i - 1 & \text{if } x = 1 - 1/2^i \\ 2i & \text{if } x = 2 - 1/2^i \end{cases}$$

Then it is easily seen that any finite increasing sequence of elements in A maps to an increasing sequence of odd integers followed by an increasing sequence of even integers. From this it follows readily that the permutations of $\text{Sub}(\pi : A \rightarrow B)$ are precisely those that consist of two increasing segments. As shown in [2] this closed set has basis $\{321, 3142, 2143\}$. Notice that A and B have order types 2ω and ω . This particular closed set cannot be defined as $\text{Sub}(\pi : A \rightarrow B)$ with both A and B having order type ω .

We now give several conditions on a closed set equivalent to it being expressible as $\text{Sub}(\pi : A \rightarrow B)$.

Theorem 1.2 *The following conditions on a closed set X are equivalent:*

1. $X = \text{Sub}(\pi : A \rightarrow B)$ for some sets A, B and bijection π .
2. X cannot be expressed as a union of two proper closed subsets.
3. For every $\alpha, \beta \in X$ there exists $\gamma \in X$ such that $\alpha \preceq \gamma$ and $\beta \preceq \gamma$.
4. X contains permutations $\gamma_1 \preceq \gamma_2 \preceq \dots$ such that, for every $\alpha \in X$, we have $\alpha \preceq \gamma_n$ for some n .

PROOF:

$1 \Rightarrow 2$. Suppose $X = \text{Sub}(\pi : A \rightarrow B)$ and yet there exist proper closed subsets Y, Z of X such that $X = Y \cup Z$. Then there exist permutations $\rho \in X \setminus Y$ and $\sigma \in X \setminus Z$. Therefore we can find subsequences $r_1 < r_2 < \dots$ and $s_1 < s_2 < \dots$ of A which are mapped by π to subsequences order isomorphic to ρ and σ . The union of $\{r_1, r_2, \dots\}$ with $\{s_1, s_2, \dots\}$ defines a sequence $t_1 < t_2 < \dots$ that is mapped by π to a subsequence order isomorphic to a permutation $\tau \in X$. Obviously, $\rho \preceq \tau$ and $\sigma \preceq \tau$. However τ belongs to at least one of Y or Z , say $\tau \in Y$. Since Y is closed we have $\rho \in Y$, a contradiction.

$2 \Rightarrow 3$. Suppose that there exist $\alpha, \beta \in X$ with the property that no permutation of X involves both of them. Put

$$\begin{aligned} Y &= \{\gamma \in X \mid \alpha \not\preceq \gamma\} \\ Z &= \{\gamma \in X \mid \beta \not\preceq \gamma\} \end{aligned}$$

Then Y and Z are proper closed subsets of X whose union is X (since any $\gamma \in X \setminus (Y \cup Z)$ would involve both α and β).

$3 \Rightarrow 4$. If θ, ϕ are two permutations in X we know that there exists a permutation of X that involves both. Temporarily we shall use the notation $\theta \vee \phi$ to denote one of these permutations. Now let $\beta_1, \beta_2, \dots$ be any listing of the permutations of X . We define a sequence of permutations of X as follows: $\gamma_1 = \beta_1$ and, for $i \geq 2$, $\gamma_i = \gamma_{i-1} \vee \beta_i$. Obviously, $\gamma_1 \preceq \gamma_2 \preceq \dots$ and, for each permutation $\beta_n \in X$, $\beta_n \preceq \gamma_n$.

$4 \Rightarrow 1$. In the sequence $\gamma_1 \preceq \gamma_2 \preceq \dots$ we remove duplicates (if any) and we insert suitable permutations so that we have one of every length. This gives a sequence of permutations $\alpha_1 \preceq \alpha_2 \preceq \dots$ such that

1. $|\alpha_i| = i$,
2. $\alpha_i \in X$,
3. for all $\sigma \in X$ there exists some α_i with $\sigma \preceq \alpha_i$

Now we shall inductively define, for each i , sets A_i, B_i and bijections $\pi_i : A_i \rightarrow B_i$ with the following properties:

1. $|A_i| = |B_i| = i$
2. π_i is order isomorphic to α_i
3. $A_{i-1} \subset A_i$ and $B_{i-1} \subset B_i$
4. $\pi_i|_{A_{i-1}} = \pi_{i-1}$

Once these sets have been constructed we can complete the proof by setting $A = \bigcup_i A_i$ and $B = \bigcup_i B_i$. Then we define $\pi : A \rightarrow B$ for any $a \in A$ by finding some A_i for which $a \in A_i$ and setting $\pi(a) = \pi_i(a)$; by the last two properties π is well-defined. The second property guarantees that $X = \text{Sub}(\pi : A \rightarrow B)$.

To carry out the construction we shall define A_i, B_i as subsets of the open interval $(0, 1)$. We begin by setting $A_1 = B_1 = \{1/2\}$ and $\pi_1(1/2) = 1/2$. Suppose now that A_i, B_i, π_i have been constructed for $i = 1, 2, \dots, n$. The permutation α_{n+1} is constructed from α_n by the insertion of a new element t at position s in α_n ; the position numbers of all the elements of α_n which are greater than or equal to s have to be increased by 1 and those values which are greater than or equal to t have also to be incremented by 1.

We reflect this insertion in the definition of A_{n+1} , B_{n+1} and π_{n+1} . The set A_{n+1} is formed by augmenting A_n with another number a whose value lies between its $(s-1)^{th}$ and s^{th} elements (if $s=1$ we take a between 0 and the minimal element of A_n ; while if $s=n$ we take a between the maximal element and 1). Similarly B_{n+1} is formed by augmenting B_n with a number b whose value lies between its $(t-1)^{th}$ and t^{th} elements. Then we define π_{n+1} so that it agrees with π_n on the elements of A_n and has $\pi_{n+1}(a) = b$. ■

Because of this result we call closed sets of the form $\text{Sub}(\pi : A \rightarrow B)$ *atomic* on the grounds that they cannot be decomposed as a proper union of two closed subsets. Expressing a given closed set as a union of atomic sets is often very useful in discovering structural information.

Example 1.3 (See [2]) $\mathcal{A}(321, 2143) = \mathcal{A}(321, 2143, 3142) \cup \mathcal{A}(321, 2143, 2413)$

Given an arbitrary closed subset one might hope to find its properties by first expressing it as a union of atomic sets, and then discovering properties of the bijection π associated with each atomic subset. Many difficulties impede this approach. It may happen that a closed set cannot be expressed as a finite union of atomic subsets. Moreover an atomic closed set may have a defining bijection π whose domain and range have high ordinal type; in that case one might be hopeful that properties of these ordinals (in particular, limit points) might imply properties of $X = \text{Sub}(\pi : A \rightarrow B)$. Despite this hope it seems sensible to begin the systematic study of atomic sets by looking at the case where the ordinal type of both A and B is that of the natural numbers $\mathbb{N}$.

2 Natural classes and sum-complete classes

A *natural class* is a closed set of the form $\text{Sub}(\pi : \mathbb{N} \rightarrow \mathbb{N})$. In other words, starting from a permutation π of the natural numbers, we form all the finite subsequences of $\pi(1), \pi(2), \dots$ and define a natural class as consisting of the permutations order isomorphic to these subsequences. From now on we shall use the notation $\text{Sub}(\pi)$ (suppressing a notational reference to the domain and range of π) in the following circumstances

1. when π is an infinite permutation with $\mathbb{N}$ as its domain and range,
2. when π is a finite permutation (in which case the domain and range are $\{1, 2, \dots, n\}$ where n is the length of π).

Example 2.1 Let π be defined by:

$$\pi = 1\ 3\ 2\ 6\ 5\ 4\ 10\ 9\ 8\ 7\ \dots$$

Then $\text{Sub}(\pi)$ is easily seen to be the set of all layered permutations as defined in [5].

If $\alpha = a_1 a_2 \cdots a_m$ and $\beta = b_1 b_2 \cdots$ are sequences (in particular, permutations) then their *sum* $\alpha \oplus \beta$ is defined to be the permutation $\gamma\delta$ where the segments γ and δ are rearrangements of $1, 2, \dots, m$ and $m+1, m+2, \dots$ respectively, and $\alpha \cong \gamma$ and $\beta \cong \delta$. Notice that we do not require that β be a finite permutation. If a permutation can be expressed as $\alpha \oplus \beta$ (with neither summand empty) we say that it is *decomposable*; otherwise it is said to be *indecomposable*. We also extend the sum notation to sets by defining, for any two sets of permutations X and Y ,

$$X \oplus Y = \{\sigma \oplus \tau \mid \sigma \in X, \tau \in Y\}$$

A set X of permutations is said to be *sum-complete* if for all $\alpha, \beta \in X$, we have $\alpha \oplus \beta \in X$. Sum-completeness and decomposability are linked by the following result, proved in [3].

Lemma 2.2 *Let X be a closed set with basis B . Then X is sum-complete if and only if B contains only indecomposable permutations.*

We shall see that natural classes and sum-complete closed sets are closely connected. The first hint of this connection is the following result which, in particular, shows that every sum-complete closed set is a natural class.

Proposition 2.3 *Let γ be any (finite) permutation and S any sum-complete closed set. Then $\text{Sub}(\gamma) \oplus S$ is a natural class.*

PROOF: Let $\beta_1, \beta_2, \dots$ be any listing of the permutations of S . Consider the sequence of permutations

$$\gamma \preceq \gamma \oplus \beta_1 \preceq \gamma \oplus \beta_1 \oplus \beta_2 \preceq \gamma \oplus \beta_1 \oplus \beta_2 \oplus \beta_3 \preceq \cdots$$

Since S is sum-complete all these permutations lie in $\text{Sub}(\gamma) \oplus S$. On the other hand it is clear that every permutation of $\text{Sub}(\gamma) \oplus S$ is involved in some term of the sequence. Hence $\text{Sub}(\gamma) \oplus S$ satisfies condition 4 of Theorem 1.2, and hence is atomic. Furthermore, the proof of (4 $\Rightarrow$ 1) in Theorem 1.2 tells us how to express $\text{Sub}(\gamma) \oplus S$ in the form $\text{Sub}(\pi : A \rightarrow B)$. Following this recipe, it is easy to see that both A and B are (isomorphic to) $\mathbb{N}$, and we have a natural class, as required. ■

Notice that the proof of this result makes no assumption on the listing of the elements of S . That means that the infinite permutation π for which $\text{Sub}(\gamma) \oplus S = \text{Sub}(\pi)$ is very far from being unique.

In the remainder of the paper we shall be exploring a partial converse of Proposition 2.3. Our main theorem will show that every finitely based natural class X does have the form of the proposition unless π and X have a very particular form.

3 A characterisation of natural classes

This section is devoted to the proof of the following theorem.

Theorem 3.1 *Let X be a finitely based natural class. Then either*

1. $X = \text{Sub}(\gamma) \oplus S$ where γ is a finite permutation and S is a sum-complete closed class determined uniquely by X , or
2. $X = \text{Sub}(\pi)$ where π is unique and ultimately periodic in the sense that there exist integers N and $P > 0$ such that, for all $n \geq N$, $\pi(n + P) = \pi(n) + P$.

The proof of the theorem will show precisely how X determines S in the first alternative. It will also, in the case of the second alternative, prove that X is enumerated by a rational generating function.

Before embarking on a series of lemmas that lead up to the proof of Theorem 3.1 we shall define some notation that will be in force for the rest of this section.

We shall let $X = \text{Sub}(\pi)$ where π is a permutation of $\mathbb{N}$. The basis of X will be denoted by B and we let b denote the length of a longest permutation in B . The permutations of B have a decomposition into sum components; the set of final components in such decompositions will be denoted by C .

We shall use the notation $\mathcal{A}(C)$ for the closed set of all permutations that avoid the permutations of C . This is a slight extension of the notation we defined in Section 1 because C might not be the basis of $\mathcal{A}(C)$ (C might contain some non-minimal elements outside $\mathcal{A}(C)$). This causes no technical difficulties. Obviously, as every permutation that avoids the permutations of C also avoids the permutations of B , we have $\mathcal{A}(C) \subseteq X$. By Lemma 2.2 $\mathcal{A}(C)$ is sum-complete; it is, as we shall see, the sum-complete class S occurring in the statement of Theorem 3.1.

From time to time we shall illustrate our proof with diagrams that display permutations. These diagrams are plots in the (x, y) plane. A permutation $p_1, p_2, \dots$ (which maps i to p_i) will be represented by a set of points whose coordinates are (i, p_i) . As a first use of such diagrams we have Figure 1 which illustrates the sum operation and the two alternatives in Theorem 3.1.

Lemma 3.2 *There exists an integer k such that, for all $d > k$,*

$$\text{Sub}(\pi(d), \pi(d+1), \dots) = \mathcal{A}(C)$$

PROOF: For each $\gamma \in C$ there is a basis element of X of the form $\beta_\gamma \oplus \gamma$. Every such β_γ is a permutation of X and so we can choose a particular subsequence $S(\beta_\gamma)$ of π with $S(\beta_\gamma) \cong \beta$. Let t be the maximal value occurring in all such $S(\beta_\gamma)$ (as γ ranges over C) and let u be the right-most position of π where an element of some $S(\beta_\gamma)$ occurs.

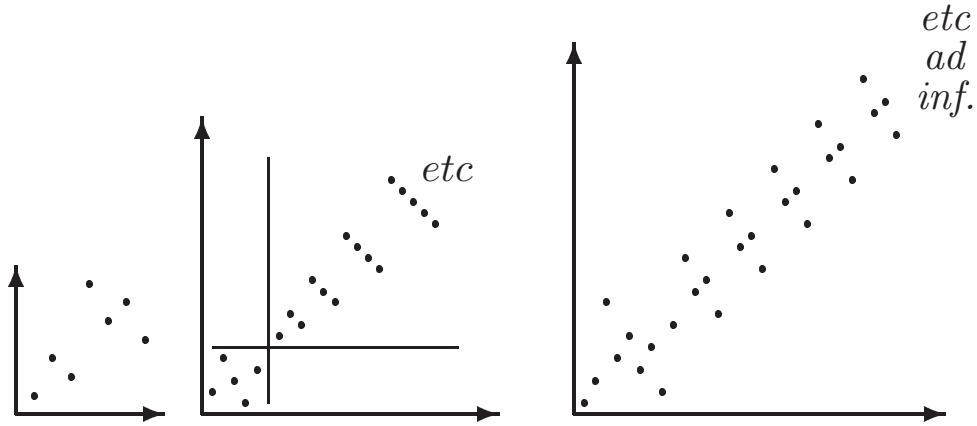


Figure 1: The sum of 132 and 4231 is 1327564, as plotted on the left. Every finitely based natural class is defined by a finite permutation summed with a sum-complete class (centre), or is eventually periodic (right).

There exists an integer $k > u$ such that all terms $\pi(k+1), \pi(k+2), \dots$ exceed t . Note that the order type of $\mathbb{N}$ is used in establishing the existence of k . Among the terms $\pi(k+1), \pi(k+2), \dots$ there can be no subsequence order isomorphic to an element of C . This proves that $\text{Sub}(\pi(d), \pi(d+1), \dots) \subseteq \mathcal{A}(C)$ for all $d > k$. It also proves that $\mathcal{A}(C)$ is non-empty.

Now let $\theta \in \mathcal{A}(C)$. Since the permutation 1 lies in $\mathcal{A}(C)$ and $\mathcal{A}(C)$ is sum-complete we have $1, 2, \dots, d-1 \oplus \theta \in \mathcal{A}(C)$. Therefore π has a subsequence order isomorphic to this permutation and that implies that $\pi(d), \pi(d+1), \dots$ has a subsequence order isomorphic to θ which completes the proof. ■

Corollary 3.3 *Either $X = \text{Sub}(\gamma) \oplus \mathcal{A}(C)$ for some finite permutation γ , or π has finitely many components and the last component (which is necessarily infinite) involves an element of C .*

PROOF: Let $\pi = \pi_1 \oplus \pi_2 \oplus \dots$ be the sum decomposition of π . Lemma 3.2 tells us, in particular, that there is a maximal position k where a subsequence order isomorphic to an element of C can begin. Suppose this position occurs in the sum component π_r . If π_r is not the final component of π then we have $\text{Sub}(\pi) = \text{Sub}(\pi_1 \oplus \dots \oplus \pi_r) \oplus \text{Sub}(\pi_{r+1} \oplus \dots)$. However $\gamma = \pi_1 \oplus \dots \oplus \pi_r$ is finite and $\text{Sub}(\pi_{r+1} \oplus \dots) = \mathcal{A}(C)$ by the lemma. ■

The first alternative of this corollary leads to the first alternative of Theorem 3.1 because of the following uniqueness result.

Proposition 3.4 *If $X = \text{Sub}(\gamma_1) \oplus S_1 = \text{Sub}(\gamma_2) \oplus S_2$ where γ_1, γ_2 are finite permutations and S_1, S_2 are sum-complete then $S_1 = S_2$.*

PROOF: Let $\sigma_1 \in S_1$. Then, as S_1 contains every permutation of the form $\iota_m = 1\ 2\ \dots\ m$, S_1 also contains $\iota_t \oplus \sigma_1$ where $t = |\gamma_2|$. But this permutation belongs to $\text{Sub}(\gamma_2) \oplus S_2$ and so can be expressed as $\gamma' \oplus \sigma_2$ where $\gamma' \preceq \gamma_2$. Since $\iota_t \oplus \sigma_1 = \gamma' \oplus \sigma_2$ and $|\gamma'| \leq |\iota_t|$ we have $\sigma_1 \preceq \sigma_2$. This proves that $\sigma_1 \in S_2$ and therefore $S_1 \subseteq S_2$. The result now follows by symmetry. ■

In the remainder of the proof of Theorem 3.1 we shall assume that the second alternative of Corollary 3.3 holds and work towards proving the second alternative of the theorem. In particular, there exists a greatest position k in π where a subsequence isomorphic to a permutation in C can begin, and this position occurs in the final (infinite) sum component π_z of π .

Next we prepare the ground for two arguments that occur later in the proof and which depend upon the indecomposability of π_z . Suppose that r is any position in π_z . We define a pair of sequences $U(r) = u_1 u_2 \dots$ and $V(r) = v_1 v_2 \dots$ by the following rules:

1. v_i is the position among the terms of π up to and including position u_{i-1} (when $i = 1$ take $u_0 = r$) where the greatest element occurs:

$$\pi(v_i) = \max\{\pi(h) \mid h \leq u_{i-1}\}.$$

2. u_i is the rightmost position in π where a term not exceeding $\pi(v_i)$ occurs:

$$u_i = \max\{h \mid \pi(h) \leq \pi(v_i)\}.$$

Figure 2 depicts these points and the next lemma assures us that the figure accurately represents the relative positions of the marked points.

Lemma 3.5 *The relative positions and sizes of the terms $\pi(u_i)$ and $\pi(v_j)$ are described by the following inequalities:*

$$\begin{aligned} v_1 &< v_2 < u_1 < v_3 < u_2 < v_4 < u_3 < \dots \\ \pi(u_1) &< \pi(v_1) < \pi(u_2) < \pi(v_2) < \pi(u_3) < \pi(v_3) < \dots \end{aligned}$$

PROOF: (I) From the definition of v_i we have $v_i \leq u_{i-1}$, and from the definition of u_i we have $\pi(v_i) \geq \pi(u_i)$. Note that we cannot have $u_i = u_{i-1}$, because then every term of π to the left of this position would be less than or equal to $\pi(v_i)$, and every term to the right would be greater than $\pi(v_i)$, contradicting the assumption that $\pi(v_i)$ belongs to the final component of π . Hence we have $v_i \leq u_{i-1} < u_i$ and $\pi(v_i) > \pi(u_i)$.

(II) By the definition of v_{i+1} , we have $v_{i+1} \leq u_i$ and $\pi(v_{i+1}) \geq \pi(v_i)$. We cannot have $v_{i+1} < v_i$, because $\pi(v_i)$ is the maximal value of π on the interval $[1, u_{i-1}]$. Also, we cannot have $v_{i+1} = v_i$, because that would imply $u_{i+1} = u_i$, which is proved impossible as

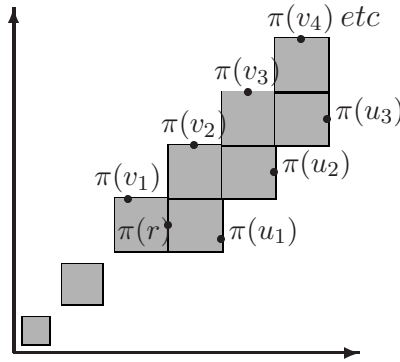


Figure 2: The terms of π as mapped out by $\pi(u_i)$ and $\pi(v_i)$. All terms lie in the shaded boxes.

in (I). Finally, we cannot have $v_{i+1} = u_i$ because $\pi(u_i) < \pi(v_i)$ by (I). We conclude that $v_i < v_{i+1} < u_i$ and $\pi(v_i) < \pi(v_{i+1})$.

(III) As in (I), we have $u_{i+1} > u_i$ and $\pi(u_{i+1}) < \pi(v_{i+1})$. Moreover, $u_{i+1} > u_i$ immediately implies that $\pi(u_{i+1}) > \pi(v_i)$.

(IV) As in (II), we have $v_{i+2} < u_{i+1}$ and $\pi(v_{i+2}) > \pi(v_{i+1})$. Moreover, $v_{i+2} > u_i$ for otherwise we would have $\pi(v_{i+2}) \leq \pi(v_{i+1})$.

Summarising (I)–(IV), we have $v_i < v_{i+1} < u_i < v_{i+2} < u_{i+1}$ and $\pi(u_i) < \pi(v_i) < \pi(u_{i+1}) < \pi(v_{i+1}) < \pi(v_{i+2})$ for every $i = 1, 2, \dots$, which is enough to prove the lemma. ■

Our first use of the sequences $U(r)$ and $V(r)$ and the above lemma occurs immediately. We have seen (Lemma 3.2) that there is a rightmost position in π where subsequences order isomorphic to permutations in C can begin. Now we prove that there is a rightmost position by which they have all ended.

Lemma 3.6 *There exists a position ℓ of π such that no subsequence of π that is order isomorphic to an element of C terminates after position ℓ .*

PROOF: Consider the sequences $U(k), V(k)$ and refer to Figure 2 with $r = k$, in particular to the edge-connected strip of boxes that begins with the box B_1 bounded by $\pi(v_1)$ and $\pi(k)$. Let S be a subsequence of π isomorphic to a permutation $\gamma \in C$. By definition of k , S cannot start to the right of $\pi(k)$. In fact, since γ is indecomposable, S must start in B_1 , and the terms of S must lie in a contiguous segment of boxes. Therefore, as $|S| \leq b$, S cannot extend beyond position $u_{\lfloor b/2 \rfloor}$. ■

In view of this lemma we may define ℓ as the last position of π that is part of a subsequence isomorphic to an element of C . Now we define the sequences $U(\ell), V(\ell)$ (and, re-using notation, call them $u_1, u_2, \dots$ and $v_1, v_2, \dots$).

The defining property of ℓ implies that $\pi(1), \dots, \pi(\ell)$ is the only subsequence of π with this order isomorphism type. For any subsequence of π order isomorphic to $\pi(1), \dots, \pi(\ell)$

has a final element that is part of a subsequence order isomorphic to an element of C . Therefore this final element cannot occur after position ℓ within π and so it must be $\pi(1), \dots, \pi(\ell)$ itself. Notice also, again from the definition of ℓ , that the permutation order isomorphic to $\pi(1), \dots, \pi(\ell)$ is the longest permutation in X whose last element is the terminating element of a subsequence order isomorphic to an element of C ; as such, this permutation depends on X and not on π . In the next lemma we prove that a number of other initial segments of π are unique of their isomorphism type, and depend on X rather than on π .

Lemma 3.7 *For each $i = 1, 2, \dots$ the sequence $\pi(1), \dots, \pi(u_i)$ is the unique subsequence of that order isomorphism type. Its corresponding permutation $\beta = \beta(1), \dots, \beta(u_i)$ is the longest permutation in X satisfying the following two properties:*

- (1) $\beta(1), \dots, \beta(u_{i-1})$ is isomorphic to $\pi(1), \dots, \pi(u_{i-1})$ (where $u_0 = \ell$);
- (2) $\beta(u_i) \leq \beta(v_i)$.

As such, β depends on X only, and not on π .

PROOF: We prove the lemma by induction, anchoring it at $u_0 = \ell$. Assume that the statements are true for some $i \geq 1$, and consider any subsequence $\pi(s_1), \dots, \pi(s_{u_{i+1}})$ order isomorphic to $\pi(1), \dots, \pi(u_{i+1})$. By the inductive hypothesis we must have $s_j = j$ for $j = 1, \dots, u_i$. But then, since $\pi(u_{i+1})$ is the rightmost term of π smaller than $\pi(v_{i+1})$, and since there are $u_{i+1} - u_i - 1$ terms between $\pi(u_i)$ and $\pi(u_{i+1})$, it follows that $s_j = j$ for $j = u_i + 1, \dots, u_{i+1}$ as well.

Clearly, the permutation β satisfies properties (1) and (2), by virtue of being isomorphic to $\pi(1), \dots, \pi(u_{i+1})$. Suppose that $\gamma = \gamma(1), \dots, \gamma(m)$ is any permutation satisfying these conditions. Consider an embedding $\pi(t_1), \dots, \pi(t_m)$ of γ in π . As above, we must have $t_j = j$ for $j = 1, \dots, u_i$. And again, $\pi(u_{i+1})$ being the rightmost term of π smaller than $\pi(v_{i+1})$, we have that $t_m \leq u_{i+1}$. But this, in turn, implies that $|\gamma| = m \leq t_m \leq u_{i+1} = |\beta|$. This proves that β is indeed the longest permutation of X satisfying (1) and (2). The last statement of the lemma is now straightforward. ■

At this point we can prove the uniqueness of π : it is the limit of its initial segments $\pi(1) \cdots \pi(u_i)$ and these depend on X alone.

For future use we record the following result, the proof of which is analogous to the proof of Lemma 3.7:

Lemma 3.8 *For each $i = 2, 3, \dots$ the subsequence consisting of all terms of π not exceeding $\pi(v_i)$ is unique of its order isomorphism type. Its corresponding permutation $\beta = \beta(1), \dots, \beta(n_i)$ is the longest permutation in X satisfying the following two properties:*

- (1) $\beta(1), \dots, \beta(u_{i-1})$ is isomorphic to $\pi(1), \dots, \pi(u_{i-1})$;
- (2) $\beta(v_i)$ is its largest term.

As such, β depends on X only, and not on π .

The subsequences whose embeddings are unique in the previous two lemmas are those all of whose terms are taken from an initial contiguous strip of blocks in Figure 2 (with $r = \ell$). For convenience we let $\sigma(i)$ be the permutation isomorphic to the first type of subsequence (Lemma 3.7), and $\sigma'(i)$ be the permutation isomorphic to the second type (Lemma 3.8).

Lemma 3.9 $u_{i+1} - u_i \leq 2(b-1)^2$

PROOF: Of course $u_{i+1} - u_i$ is the number of terms of $\pi(u_i + 1) \cdots \pi(u_{i+1})$. We divide these terms into two sets

$$L = \{j \mid u_i < j \leq u_{i+1} \text{ and } \pi(j) < \pi(v_{i+1})\}$$

and

$$U = \{j \mid u_i < j \leq u_{i+1} \text{ and } \pi(j) > \pi(v_{i+1})\}$$

and we shall show that both $|L|$ and $|U|$ are at most $(b-1)^2$. Each bound is proved in the same way and we give the details for $|L|$ only. Figure 3 depicts the locations of L and U within π .

Consider a maximal increasing subsequence $\pi(j_1)\pi(j_2) \cdots \pi(j_m)$ of π such that all $j_i \in L$. Using this we form another subsequence of π whose terms are the following:

1. all terms not exceeding $\pi(v_i)$, (a subsequence order isomorphic to $\sigma'(i)$)
2. the term $\pi(v_{i+1})$
3. all the terms $\pi(j_1)\pi(j_2) \cdots \pi(j_m)$

The permutation which is order isomorphic to this subsequence is, of course, a member of X and we write it as $\lambda n \rho \theta$ where n corresponds to the term $\pi(v_{i+1})$, $\lambda \rho$ corresponds to $\sigma'(i)$, and $\theta = n - m, n - m + 1, \dots, n - 1$ corresponds to $\pi(j_1)\pi(j_2) \cdots \pi(j_m)$.

Now consider another permutation almost the same as this except that θ contains one more term, and n is replaced by $n + 1$. It has the form $\lambda, n + 1, \rho \theta'$ with $\theta' = n - m, n - m + 1, \dots, n - 1, n$. This permutation does not belong to X . To see this, assume that some subsequence of π is order isomorphic to it. In the correspondence between the permutation and the subsequence, $\lambda \rho$ (which is order isomorphic to $\sigma'(i)$) must be mapped to the subsequence of terms not exceeding $\pi(v_i)$, by Lemma 3.8, and $n + 1$ must be mapped to one of the terms of π in the range of positions $u_{i-1} + 1$ to u_i . This forces

θ' to be mapped into L as these are the only positions of π to the right of u_i and smaller than $\pi(v_{i+1})$. However, this contradicts that L contains no increasing sequence of length $m + 1$.

It follows that $\lambda, n + 1, \rho\theta'$ must involve a basis element of X . A particular embedding of a basis element must contain *all* the terms of θ' for otherwise this basis element would be embedded in $\lambda n \rho \theta$ which is impossible. In particular we can deduce that $m + 1 \leq b$.

Exactly the same argument can be carried out for maximal decreasing subsequences. Thus the sequence $\pi(L)$ contains no increasing or decreasing subsequence of length more than $b - 1$ and, by the well known result of Erdős and Szekeres (see [6]), we conclude that $|L| \leq (b - 1)^2$.

The proof that $|U| \leq (b - 1)^2$ is similar but it uses $\pi(u_{i+1})$ and $\sigma(i)$ instead of $\pi(v_{i+1})$ and $\sigma'(i)$. ■

We now define an encoding of permutations of X . If $\xi = x_1 x_2 \cdots x_n \in X$ then we encode it as $E(\xi) = e_1 e_2 \cdots e_n$ where

$$e_k = |\{i \mid 1 \leq i \leq k \text{ and } x_i \geq x_k\}|$$

If a term of π lies in the final component beyond position u_2 then there are, by Lemma 3.9, at most $4(b - 1)^2$ preceding terms greater than it. On the other hand, if it lies in one of the finite components or in the final component and not beyond position u_2 then, obviously, there will again only be a bounded number of preceding greater terms. Since permutations of X are order isomorphic to subsequences of π and π is determined by X there is an upper bound depending on X alone for each of the code symbols e_k . Thus we may consider $E(X)$ as a language over some finite alphabet $A = \{1, \dots, m\}$.

The encoding has the property that, if $\gamma = \delta\epsilon$ is a permutation and $\bar{\delta}$ is the permutation order isomorphic to the initial segment δ , then $E(\bar{\delta})$ is an initial segment of $E(\gamma)$.

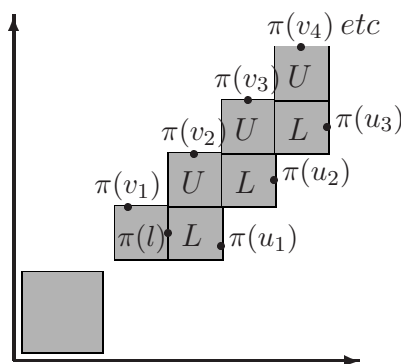


Figure 3: π mapped out by u_i and v_i . None of the boxes marked L or U contains a monotonic increasing or decreasing subsequence of length b . Thus the sizes of these boxes is bounded.

Furthermore this encoding (in which every element of a permutation is encoded by the number of its higher predecessors) is closely related to the encoding studied in [1] (in which every element was encoded by its number of lower successors). In fact, if $F(\xi)$ denotes the latter encoding and $\bar{\xi}$ is the permutation obtained from ξ by replacing each element x_i by $|\xi| - x_i + 1$ and then reversing it, then $F(\bar{\xi})$ is the reverse of $E(\xi)$. It therefore follows from Theorem 2 of [1] that

Lemma 3.10 *$E(X)$ is a regular set.*

We can now confirm one of the claims we made when stating Theorem 3.1: from the results of [1] the ordinary generating function of the sequence (g_n) , where g_n is the number of permutations of length n , is a rational function. But to show that π is eventually periodic and thus complete the proof of the theorem we need to study a deterministic finite automaton that accepts $E(X)$. We denote this automaton by $\mathcal{M} = (\Sigma, A, s_0, \tau, T)$ (the notation specifies, respectively, the set of states, the alphabet, the initial state, the transition function, and the set of final states).

Lemma 3.11 *$\{E(\sigma(i)) \mid i = 1, 2, \dots\}$ contains the set*

$$\{\alpha\beta^j \mid j = 0, 1, 2, \dots\}$$

defined by the regular expression $\alpha\beta^$ for some non-empty words $\alpha, \beta \in A^*$.*

PROOF: We shall consider the sequence of states $t_i = \tau(s_0, E(\sigma(i)))$, $i = 1, 2, \dots$ and aim to show that it is periodic. By definition and by Lemma 3.7 $\sigma(i+1) = \theta\phi$ where

- (a) θ is a sequence order isomorphic to $\sigma(i)$
- (b) $\phi = a_1a_2 \cdots a_n$ satisfies $a_n < \max(\theta)$ and $\theta\phi \in X$
- (c) ϕ is maximal with these properties.

We shall express these conditions in terms of the automaton $\mathcal{M}$. From (a), (b) and the definition of E we have

$$E(\sigma(i+1)) = E(\sigma(i))b_1b_2 \cdots b_n$$

where $w = b_1b_2 \cdots b_n$ is a word in the alphabet A and $\tau(t_i, w) = t_{i+1}$.

By definition, b_j is the number of terms of $\theta\phi$ up to and including a_j that exceed or equal a_j . To capture the condition $a_n < \max(\theta)$ we need to define another sequence $c_0c_1c_2 \cdots c_n$ where $c_0 = 0$ and c_j is the number of terms up to and including a_j that exceed $\max(\theta)$; of course, all such terms are among $a_1, a_2, \dots, a_n$.

If $a_j < \max(\theta)$ then the terms enumerated by b_j include $\max(\theta)$, a_j and the c_{j-1} terms above $\max(\theta)$; hence $b_j > c_{j-1} + 1$. However, if $a_j > \max(\theta)$ then each of the b_j terms that exceed or equal a_j is one of the terms that exceeds $\max(\theta)$; hence $b_j \leq c_{j-1} + 1$.

Furthermore, in the former case $c_j = c_{j-1}$ and in the latter case $c_j = c_{j-1} + 1$. Thus $c_1 c_2 \cdots c_n$ is determined uniquely by $b_1 b_2 \cdots b_n$, and $a_n < \max(\theta)$ if and only if $b_n > c_n$.

Putting all this together t_{i+1} is the unique state of $\mathcal{M}$ for which there exists a word $w = b_1 b_2 \cdots b_n$ in the alphabet A with the following properties:

1. $\tau(t_i, w) = t_{i+1}$;
2. if the sequence $(c_0, c_1, \dots, c_n)$ is defined by $c_0 = 0$ and, for $j > 0$,

$$c_j = \begin{cases} c_{j-1} & \text{if } b_j > c_{j-1} + 1, \\ c_{j-1} + 1 & \text{if } b_j \leq c_{j-1} + 1 \end{cases}$$

then $b_n > c_n$;

3. w has maximal length among all words satisfying these two conditions.

But now note that the three conditions depend only on the t_i and the automaton $\mathcal{M}$ and not on $E(\sigma(i))$. Therefore the sequence $t_1, t_2, \dots$ is ultimately periodic. So, for some $P > 0$ and N we have $t_j = t_{j+P}$ for all $j \geq N$.

Let $\alpha = E(\sigma(N))$ and let β be the unique word such that $E(\sigma(N + P)) = \alpha\beta$. Then $E(\sigma(N + hP)) = \alpha\beta^h$ and $\tau(s_0, \alpha\beta^h) = t_N$ for all $h \geq 0$. This proves that $\alpha\beta^* \subseteq \{E(\sigma(i)) \mid i = 1, 2, \dots\}$. ■

We can now complete the proof of Theorem 3.1. In the notation of the previous lemma, let $|\alpha| = m$ and $|\beta| = n$. Consider the encoding of π itself: $E(\pi) = e_1 e_2 \cdots$. This is just the limit of its prefixes $E(\sigma(1)), E(\sigma(2)), \dots$. It is also the limit of $E(\sigma(N)), E(\sigma(N + P)), E(\sigma(N + 2P)), \dots$. Hence, by Lemma 3.11, $E(\pi)$ is ultimately periodic with $e_{j+n} = e_j$ for all $j \geq m$.

Consider an arbitrary $\pi(j)$ with $j \geq m$. We have $\pi(j) = l_j + r_j + 1$ where

$$l_j = |\{i \mid i \leq j \text{ and } \pi(i) < \pi(j)\}|$$

and

$$r_j = |\{i \mid i > j \text{ and } \pi(i) < \pi(j)\}|$$

Obviously, $l_j = j - e_j$. The number r_j can be obtained from $E(\pi)$ as follows. Define two sequences $\nu^{(j)} = (n_0, n_1, \dots)$ and $\theta^{(j)} = (h_0, h_1, \dots)$ by $n_0 = 0$ and $h_0 = e_j$ and

$$n_{i+1} = \begin{cases} n_i + 1 & \text{if } e_{j+i+1} > h_i \\ n_i & \text{otherwise} \end{cases}$$

and

$$h_{i+1} = \begin{cases} h_i & \text{if } e_{j+i+1} > h_i \\ h_i + 1 & \text{otherwise} \end{cases}$$

An easy inductive argument shows that n_i is equal to the number of terms from $\pi(j+1), \dots, \pi(j+i)$ which are smaller than $\pi(j)$, while $h_i - e_j$ is the number of terms from the same set which are greater than $\pi(j)$. In particular, $\nu^{(j)}$ eventually becomes constant with value r_j .

Finally, note that $\nu^{(j)}$ depends only on $e_j e_{j+1} e_{j+2} \dots$ and not on $e_1 e_2 \dots e_{j-1}$. Hence $\nu^{(j)} = \nu^{(j+n)}$ and $r_j = r_{j+n}$. Therefore

$$\begin{aligned} \pi(j+n) &= l_{j+n} + r_{j+n} + 1 \\ &= j + n - e_{j+n} + r_{j+n} + 1 \\ &= n + j - e_j + r_j + 1 \\ &= n + l_j + r_j + 1 \\ &= n + \pi(j) \end{aligned}$$

as required.

4 Natural classes with infinite bases

Any natural class that is not of the form stipulated by the conclusion of Theorem 3.1 is, of course, not finitely based. An example of such a class is $Y = \text{Sub}(\pi)$ where

$$\pi = 3 \ 2 \ 5 \ 1 \ [7, 8] \ 4 \ [10, 12] \ 6 \ [14, 17] \ 9 \ [19, 23] \ 13 \ [25, 30] \ 18 \ [32, 38] \ 24 \dots$$

In this example $[a, b]$ stands for the segment $[a, a+1, \dots, b]$. It is clear by inspection that π is not periodic and so, by the uniqueness conclusion of Theorem 3.1, Y is not of periodic type. We argue that it is not of the form $\text{Sub}(\gamma) \oplus S$ where S is sum-complete. Suppose it were of this form. Consider the initial segments of π ending with one of $1, 4, 6, 9, \dots$ respectively. These all define permutations $\xi_1, \xi_2, \xi_3, \xi_4, \dots$ of Y . Every ξ_i is indecomposable and has a unique embedding in π . From the indecomposability those ξ_i of length greater than γ must be order isomorphic to permutations of S ; but, if $\xi_i \in S$, so also is $\xi_i \oplus \xi_i$ which contradicts that it is uniquely embeddable in π .

It would perhaps be tempting to suppose that when π is periodic, the closed class $\text{Sub}(\pi)$ is always finitely based. This, however, is not the case, as our final example shows.

Let $X = \text{Sub}(\pi)$, where

$$\pi = \underline{2} \ \underline{3} \ \underline{5} \ 1 \ \underline{7} \ \underline{8} \ 4 \ \underline{10} \ 6 \ \underline{12} \ \underline{13} \ 9 \ \underline{15} \ 11 \dots$$

Essentially, π is an increasing oscillating sequence with every other left maximal term replaced with an increasing pair (the underlinings are intended to highlight this). Call these increasing pairs *twins*, and note that they are the only pairs of terms of π occurring in successive positions and having successive values.

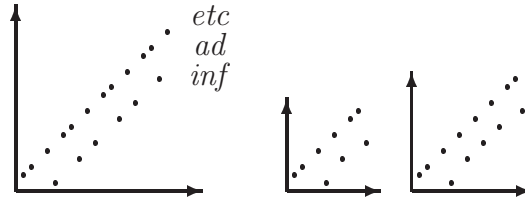


Figure 4: On the left, an infinitely based periodic natural class. On the right, two basis elements.

We claim that each of the following permutations belongs to the basis of X :

$$\begin{aligned}
 \beta_1 &= \underline{2} \underline{3} \underline{4} \underline{5} 1 \\
 \beta_2 &= \underline{2} \underline{3} \underline{5} 1 \underline{7} 4 \underline{8} \underline{9} 6 \\
 &\vdots \\
 \beta_n &= \underline{2} \underline{3} \underline{5} 1 \underline{7} 4 \underline{9} 6 \cdots \underline{4n-3} 4n-6 \underline{4n-1} 4n-4 \underline{4n} \underline{4n+1} 4n-2 \\
 &\vdots
 \end{aligned}$$

The permutation β_n is obtained from an oscillating sequence with an even number of left maximal terms by replacing the first and last of these terms by increasing pairs (see Figure 4). To show that $\beta_n \notin X$ we can argue as follows. Write π as

$$\pi = U_0 U_1 l_1 U_2 l_2 U_3 l_3 \dots,$$

where $U_0, U_2, U_4, \dots$ are the twins, $U_1, U_3, U_5, \dots$ are the remaining (unexpanded) left maxima, and $l_1, l_2, l_3, \dots$ are the remaining terms. Suppose β_n embeds into π . The two twins of β_n must correspond to two twins, say U_{2p} and U_{2q} , of π . Since β_n is indecomposable, the remaining left maxima $5, 7, 9, \dots$ of β_n must map into $U_{2p+1}, U_{2p+2}, U_{2p+3}, \dots$ respectively. The number of left maxima between the two twins of β_n is even, while the number of segments $U_{2p+1}, \dots, U_{2q-1}$ is odd, a contradiction.

To complete the proof that β_n is a basis permutation of X , we need to demonstrate that $\beta_n \setminus \{\beta_n(j)\}$, the permutation obtained by removing the j th term from β_n , belongs to X for every $j = 1, \dots, 4n+1$. If $j \notin \{1, 2, 4n-1, 4n\}$ the resulting permutation is decomposable, and can be embedded into π by embedding each of its components and keeping them sufficiently apart. If j is one of $1, 2, 4n-1$ or $4n$ then one of the twins of β_n becomes a singleton. Suppose, for the sake of argument, that $j = 4n$ (the other cases are treated analogously). Then we can embed $\beta_n \setminus \{\beta_n(4n)\}$ by mapping 2 and 3 onto U_0 , all the other left maxima of β_n into $U_1, U_2, \dots$ respectively, and the remaining terms into $l_1, l_2, \dots$ respectively. Note that the parity problem which prevented us from embedding β_n into π does not arise here, because the second twin of β_n has become a singleton in $\beta_n \setminus \{\beta_n(4n)\}$, and can therefore be mapped onto the singleton U_{2n-1} .

References

- [1] M. H. Albert, M. D. Atkinson, N. Ruškuc: Regular Closed Classes of Permutations, Theoretical Computer Science 306 (2003), 85–100.
- [2] M. D. Atkinson: Restricted permutations, Discrete Math. 195 (1999), 27–38.
- [3] M. D. Atkinson, T. Stitt: Restricted permutations and the wreath product, Discrete Math. 259 (2002), 19–36.
- [4] M. D. Atkinson, M. M. Murphy, N. Ruškuc: Partially well ordered closed sets of permutations, Order 19 (2002), 101–113.
- [5] M. Bóna: The Solution of a Conjecture of Wilf and Stanley for all layered patterns, Journal of Combinatorial Theory, Series A 85 (1999), 96–104.
- [6] P. Cameron: “Combinatorics”, Cambridge University Press, First Edition (1994).
- [7] M. M. Murphy: “Restricted Permutations, Antichains, Atomic Classes and Stack Sorting”, Ph.D. Thesis, University of St Andrews, St Andrews, Scotland, UK (2002).
- [8] N. Ray, J. West: Posets of matrices and permutations with forbidden subsequences, Annals of Combinatorics 7 (2003), 55–88.
- [9] D. A. Spielman, M. Bóna: An infinite antichain of permutations, Electronic J. Combinat. 7 (2000), #N2.