The Dyck pattern poset

Axel Bacher, Antonio Bernini, Luca Ferrari, Benjamin Gunby, Renzo Pinzani, Julian West

The Dyck pattern poset*

Axel Bacher[†] Antonio Bernini[‡] Luca Ferrari[‡] Benjamin Gunby[§]
Renzo Pinzani[‡] Julian West[¶]

Abstract

We introduce the notion of *pattern* in the context of lattice paths, and investigate it in the specific case of Dyck paths. Similarly to the case of permutations, the pattern-containment relation defines a poset structure on the set of all Dyck paths, which we call the *Dyck pattern poset*. Given a Dyck path P, we determine a formula for the number of Dyck paths covered by P, as well as for the number of Dyck paths covering P. We then address some typical pattern-avoidance issues, enumerating some classes of pattern-avoiding Dyck paths. We also compute the generating function of Dyck paths avoiding any single pattern in a recursive fashion, from which we deduce the exact enumeration of such a class of paths. Finally, we describe the asymptotic behavior of the sequence counting Dyck paths avoiding a generic pattern, we prove that the Dyck pattern poset is a well-ordering and we propose a list of open problems.

Proposed running head: The Dyck pattern poset. *AMS Classification (2010)*: 05A15, 05A16, 06A07.
Keywords: Dyck path, pattern, exact enumeration, asymptotic, poset.

1 Introduction

One of the most investigated and fruitful notions in contemporary combinatorics is that of a *pattern*. Historically it was first considered for permutations [11], then analogous definitions were provided in the context of many other structures, such as set partitions [8, 10, 13], words [2, 3], and trees [5, 7, 12]. Perhaps all of these examples have been motivated or informed by the more classical notion of graphs and subgraphs. Informally speaking, given a specific class of combinatorial objects, a pattern can be thought of as an occurrence of a small object inside a larger one; the word "inside" means that the pattern is suitably embedded into the larger object, depending on the specific combinatorial class of objects. The main aim of the present work is to introduce the notion of pattern in the context of lattice paths and to begin its systematic study in the special case of Dyck paths.

For our purposes, a *lattice path* is a path in the discrete plane starting at the origin of a fixed Cartesian coordinate system, ending somewhere on the x-axis, never going below the x-axis and

*Ax. Ba. supported by FWF project F050-04; An. Be. and L. F. partially supported by PRIN project: "Automi e Linguaggi Formali: Aspetti Matematici e Applicativi"; L. F. and R. P. partially supported by INdAM project: "Proprietà algebriche e combinatorie dei cammini discreti"; B. G. supported by NSF/DMS grant 1062709 and NSA grant H98230-11-1-0224.

[†]Research Institute for Symbolic Computation (RISC), Johannes Kepler University, Altenbergerstraße 69, A-4040 Linz, Austria, axel.bacher@risc.jku.at

[‡]Dipartimento di Matematica e Informatica "U. Dini", viale Morgagni 65, University of Firenze, Firenze, Italy, antonio.bernini@unifi.it, luca.ferrari@unifi.it, pinzani@dsi.unifi.it

[§]Massachusetts Institute of Technology, Department of Mathematics, bgunby314@gmail.com

[¶]Heilbronn Institute for Mathematical Research, University of Bristol, Bristol, UK, aeijlnstuw@gmail.com

using only a prescribed set of steps Γ. We will refer to such paths as Γ-*paths*. This definition is extremely restrictive if compared to what is called a lattice path in the literature, but it will be enough for our purposes. Observe that a Γ-path can be alternatively described as a finite word on the alphabet Γ obeying certain conditions. Using this language, we say that the *length* of a Γ-path is simply the length of the word which encodes such a path. Among the classical classes of lattice paths, the most common are those using only steps $U(p) = (1, 1)$, $D(own) = (1, -1)$ and $H(orizontal) = (1, 0)$; with these definitions, Dyck, Motzkin and Schröder paths correspond respectively to the set of steps $\{U, D\}$, $\{U, H, D\}$ and $\{U, H^2, D\}$.

Consider the class $\mathcal{P}_\Gamma$ of all Γ-paths, for some choice of the set of steps Γ. Given $P, Q \in \mathcal{P}_\Gamma$ having length k and n respectively, we say that Q *contains (an occurrence of) the pattern* P whenever P occurs as a (non-contiguous) subword of Q. So, for instance, in the class of Dyck paths, $UUDUDDUDUUDD$ contains the pattern $UUDDUD$, whereas in the class of Motzkin paths, $UUHDUUDHDDUDHUD$ contains the pattern $UHUDDHUD$. When Q does not contain any occurrence of P we will say that Q *avoids* P. In the Dyck case, the previously considered path $UUDUDDUDUUDD$ avoids the pattern $UUUUDDDD$.

This notion of pattern gives rise to a partial order in a very natural way, by declaring $P \leq Q$ when P occurs as a pattern in Q. In the case of Dyck paths, the resulting poset will be denoted by $\mathcal{D}$. It is immediate to notice that $\mathcal{D}$ has a minimum (the empty path), does not have a maximum, is locally finite and is ranked (the rank of a Dyck path is given by its semilength). As an example, in Figure 1 we provide the Hasse diagram of an interval in the Dyck pattern poset.

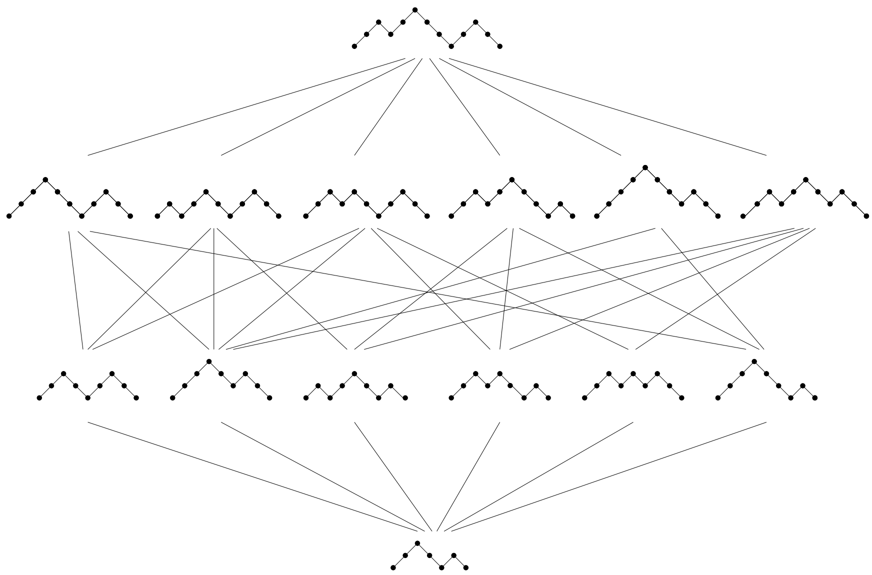

Figure 1: An interval of rank 3 in the Dyck pattern poset.

Observe that this notion of pattern for paths is very close to the analogous notion for words (considered, for instance, by Björner in [2], where the author determines the Möbius function of the associated pattern poset). Formally, instead of considering the set of *all* words of the alphabet $\{U, D\}$, we restrict ourselves to the set of Dyck words (so what we actually do is to consider a subposet of Björner's poset). However, the conditions a word has to obey in order to belong to this subposet (which translate into the fact of being a Dyck word) make this subposet highly nontrivial, and fully justify our approach, consisting of the study of its properties independently of its relationship with the full word pattern poset.

Our paper is organized as follows. In Section 2 we give two explicit formulas for the number of paths covering/covered by a given Dyck path in the Dyck pattern poset. Sections 3 and 4 are devoted to the exact enumeration of some classes of Dyck paths avoiding a single pattern. Our starting results concerns certain types of arbitrarily long patterns; we then find a set of recursive formulas which allows to automatically determine a closed form for the number of Dyck paths of a given length avoiding a single Dyck path. Further results not specifically focused on exact enumeration are contained in Section 5, namely we describe the asymptotic behavior of the number of Dyck paths avoiding a single pattern. Finally, Section 6 contains the proof that in the Dyck pattern poset there are no infinite antichains, as well as some conjectures, both enumerative and structural, suggesting that the Dyck pattern poset deserves to be better investigated.

Part of the results of the present paper (some of them without proofs) has appeared in the proceedings of the conference FPSAC 2013 [1].

2 The covering relation in the Dyck pattern poset

In the Dyck pattern poset $\mathcal{D}$, following the usual notation for covering relation, we write $P \prec Q$ (Q *covers* P) to indicate that $P \leq Q$ and the rank of P is one less than the rank of Q (i.e., $\mathrm{rank}(P) = \mathrm{rank}(Q) - 1$). Our first result concerns the enumeration of Dyck paths covered by a given Dyck path Q. We need some notation before stating it. Let $k+1$ be the number of points of Q lying on the x-axis (call such points $p_0, p_1, \ldots, p_k$). Then Q can be factorized into k Dyck factors $F_1, \ldots, F_k$, each F_i starting at p_{i-1} and ending at p_i. Let n_i be the number of ascents in F_i (an ascent being a consecutive run of U steps; n_i also counts both the number of descents and the number of peaks in F_i, where a peak in a Dyck path consists of a U step immediately followed by a D step). Moreover, we denote by $p(Q)$ and $v(Q)$ the number of occurrences in Q of a consecutive factor UDU and DUD, respectively. In the path Q of Figure 2, we have $n_1 = 2$, $n_2 = 1$, $n_3 = 2$, $p(Q) = 3$, and $v(Q) = 2$.

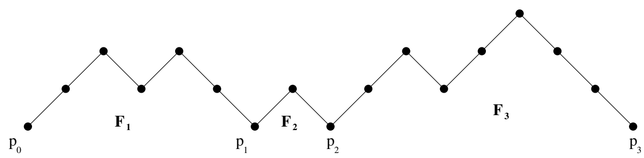

Figure 2: A Dyck path having three factors.

Proposition 2.1 *If Q is a Dyck path with k factors $F_1, \ldots, F_k$, with F_i having n_i ascents, then the number of Dyck paths covered by Q is given by*

$$\sum_{1 \leq i \leq j \leq k} n_i n_j - p(Q) - v(Q) \quad . \tag{1}$$

Proof. We proceed by induction on k. If Q is any Dyck path having only one factor (and so necessarily n_1 ascents), then a path P such that $P \prec Q$ is obtained by choosing (and then removing) a U step and a D step from an ascent and a descent of Q, respectively. This can be done in n_1^2 different ways. Note that the path P does not depend on which U from a given ascent is chosen and which D from a given descent is chosen. Moreover, for each UDU occurring in Q, removing the D step from the UDU and a U step from the ascent either immediately before D

or immediately after D produces the same path P covered by Q. An analogous argument can be used with the pattern DUD in place of UDU. Therefore, these paths would be counted twice if the term n_1^2 were not corrected by subtracting both $p(Q)$ and $v(Q)$. This leads to formula (1) in the case $k = 1$.

Now suppose that $\tilde{Q}$ is a Dyck path which has $k > 1$ factors $F_1, \dots, F_k$, each factor F_i having n_i ascents. Let l be the total number of UDU and DUD (i.e. $l = p(\tilde{Q}) + v(\tilde{Q})$) in $\tilde{Q}$. If a new factor F_{k+1} having n_{k+1} ascents and a total number l_{k+1} of UDU and DUD factors is appended to $\tilde{Q}$ (after F_k), then the paths covered by the new path Q can be obtained by removing a D step and a U step either both belonging to $\tilde{Q}$, or both belonging to F_{k+1}, or one belonging to $\tilde{Q}$ and the other one belonging to F_{k+1}.

We start by supposing that the two factors F_k and F_{k+1} are both different from UD. In the first of the above cases, the number of covered paths is given by formula (1) thanks to our inductive hypothesis (since the removal of the steps U and D involves only the first k factors of the Dyck path). The second case is easily dealt with using the induction hypothesis as well, namely applying the base case ($k = 1$) to the last factor F_{k+1}. Finally, concerning the last case, notice that the step D must be removed from $\tilde{Q}$, and the step U must be removed from F_{k+1}, otherwise the resulting path would fall below the x-axis. Then, the D step can be selected from $\sum_{i=1}^{k} n_i$ different descents of $\tilde{Q}$, while the U step can be chosen among the steps of the n_{k+1} ascents of F_{k+1}, leading to $n_{k+1} \cdot \sum_{i=1}^{k} n_i$ different paths covered by Q. Summing the contributions of the three cases considered above, we obtain:

$$\begin{aligned} & \sum_{1 \le i \le j \le k} n_i n_j - l + n_{k+1}^2 - l_{k+1} + n_{k+1} \sum_{i=1}^{k} n_i \\ = & \sum_{1 \le i \le j \le k+1} n_i n_j - l - l_{k+1} \quad . \end{aligned} \tag{2}$$

However, we still have to take into account the cases in which F_k and/or F_{k+1} are equal to UD. If $F_k = F_{k+1} = UD$, then in formula (2) we have to subtract 2 (since we have one more factor UDU and one more factor DUD than those previously counted). In the remaining cases, there is only one more factor (either UDU or DUD), thus in formula (2) we have to subtract 1. In all cases, what we get is precisely formula (1). ■

In a similar fashion, we are also able to find a formula for the number of all Dyck paths which cover a given path.

Proposition 2.2 *If P is a Dyck path of semilength n with k factors $F_1, \dots, F_k$, with F_i having semilength f_i, then the number of Dyck paths covering P is given by*

$$1 + \sum_{i \le j} f_i f_j. \tag{3}$$

Proof. A path Q covers P if and only if it is obtained from P by suitably inserting an up step U and a down step D. Thus the set of all Dyck paths covering P can be determined by choosing, in all possible ways, two positions (inside P) in which to insert an up step and a down step. Clearly, in performing these insertions, we must take care not to fall below the x-axis.

Given a Dyck path P, construct each Dyck path Q covering P by adding a U step and a D step in the rightmost possible places. Concerning the U step, we can place it either at the end of P or before one of the D steps of P. If we place it at the end of P, then necessarily also the D step must be placed at the end (after the added U step), which generates only one path Q covering P. Otherwise, suppose the U step is placed into the i-th factor of P. Then the D

step must be placed either at the end of P or before a U step in the j-th factor, with $j \geq i$, except of course before the first U step of the i-th factor of P (otherwise we would not get a Dyck path). Therefore the total number of possibilities is

$$1 + \sum_{i \leq j} f_i f_j,$$

as desired. ∎

3 Enumerative results on pattern avoiding Dyck paths

In the present section we will be concerned with the enumeration of some classes of pattern avoiding Dyck paths. Similarly to what has been done for other combinatorial structures, we are going to consider classes of Dyck paths avoiding a single pattern, and we will start examining the cases of simple patterns. Specifically, we will count Dyck paths avoiding any single path of semilength ≤ 3; each case will arise as a special case of a more general result concerning a certain class of patterns, except for the pattern $UDUUDD$, whose enumerative properties immediately follows from the study of its mirror image $UUDDUD$.

Given a pattern P, we denote by $\mathcal{D}_n(P)$ the set of all Dyck paths of semilength n avoiding the pattern P, and by $d_n(P)$ the cardinality of $\mathcal{D}_n(P)$.

3.1 The pattern $(UD)^k$

This is one of the easiest cases.

Proposition 3.1 *For any $k \in \mathbf{N}$, $Q \in \mathcal{D}_n((UD)^k)$ if and only if Q has at most $k-1$ peaks.*

Proof. For any positive a_i's and b_i's, a Dyck path $Q = U^{a_1}D^{b_1}U^{a_2}D^{b_2}\cdots U^{a_h}D^{b_h}$ contains the pattern $(UD)^k$ if and only if $h \geq k$, that is Q has at least k peaks. ∎

Since it is well known that the number of Dyck paths of semilength n and having k peaks is given by the Narayana number $N_{n,k}$ (sequence A001263 in [14]), we have that $d_n((UD)^k) = \sum_{i=0}^{k-1} N_{n,i}$ (partial sums of Narayana numbers). Thus, in particular:

- $d_n(UD) = 0$;
- $d_n(UDUD) = 1$;
- $d_n(UDUDUD) = 1 + \binom{n}{2}$.

3.2 The pattern $U^{k-1}DUD^{k-1}$

Let Q be a Dyck path of length $2n$ and $P = U^{k-1}DUD^{k-1}$. Clearly if $n < k$, then Q avoids P, and if $n = k$, then all Dyck paths of length $2n$ except one (Q itself) avoid Q. Therefore:

- $d_n(P) = C_n$ if $n < k$, and
- $d_n(P) = C_n - 1$ if $n = k$,

where C_n is the n-th Catalan number.

Now suppose $n > k$. Denote by A the end point of the $(k-1)$-th U step of Q. It is easy to verify that A belongs to the line r having equation $y = -x + 2k - 2$. Denote with B the starting point of the $(k-1)$-th-to-last D step of Q. An analogous computation shows that B belongs to the line s having equation $y = x - (2n - 2k + 2)$.

Depending on how the two lines r and s intersect, it is convenient to distinguish two cases.

Lemma 3.1 *If $n \geq 2k-3$, then*

$$d_n(P) = C_{k-1}^2. \tag{4}$$

Proof. If $2n - 2k + 2 \geq 2k - 4$ (i.e. $n \geq 2k - 3$), then r and s intersect at height ≤ 1, hence $x_A \leq x_B$ (where x_A and x_B denote the abscissas of A and B, respectively). The path Q can be split into three parts (see Figure 3): a prefix Q_A from the origin $(0,0)$ to A, a path X from A to B, and a suffix Q_B from B to the last point $(2n, 0)$.

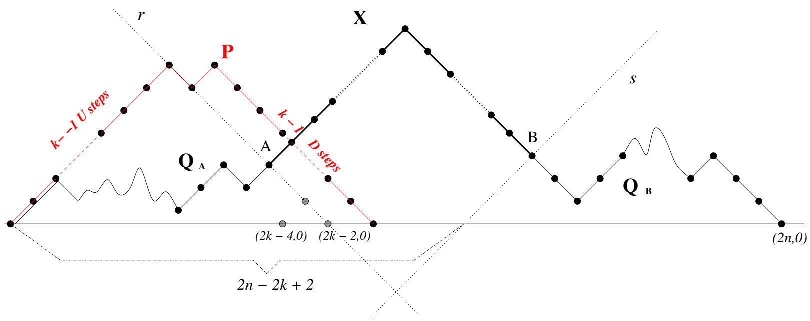

Figure 3: Avoiding $U^{k-1}DUD^{k-1}$, with $n \geq 2k-3$

We point out that Q_A has exactly $k-1$ U steps and its last step is a U step. Analogously, Q_B has exactly $k-1$ D steps and its first step is a D step. Notice that there is a clear bijection between the set $\mathcal{A}$ of Dyck prefixes having $k-1$ U steps and ending with a U and the set $\mathcal{B}$ of Dyck suffixes having $k-1$ D steps and starting with a D, since each element of $\mathcal{B}$ can be read from right to left thus obtaining an element of $\mathcal{A}$. Moreover, $\mathcal{A}$ is in bijection with the set of Dyck paths of semilength $k-1$ (just complete each element of $\mathcal{A}$ with the correct sequence of D steps), hence $|\mathcal{A}| = C_{k-1}$.

If we require Q to avoid P, then necessarily $X = U^i D^j$, for suitable i, j (for, if a valley DU occurred in X, then Q would contain P since U^{k-1} and D^{k-1} already occur in Q_A and Q_B, respectively). In other words, A and B can be connected only in one way, using a certain number (possibly zero) of U steps followed by a certain number (possibly zero) of D steps. Therefore, a path Q avoiding P is essentially constructed by choosing a prefix Q_A from $\mathcal{A}$ and a suffix Q_B from $\mathcal{B}$, hence $d_n(P) = C_{k-1}^2$, as desired. ■

Lemma 3.2 *If $k+1 \leq n < 2k-3$, then*

$$d_n(P) = \left(\sum_{j=0}^{n-k+1} b_{k-2,j} \right)^2 + \sum_{j \geq 2} b_{k-j,n-k+j}^2 \tag{5}$$

Proof. If $k+1 \leq n < 2k-3$, then r and s intersect at height > 1. It can be either $x_A \leq x_B$ or $x_A > x_B$.

a) If $x_A \leq x_B$, then we can count all Dyck paths Q avoiding P using an argument analogous to the previous lemma. However, in this case the set of allowable prefixes of each such Q is a proper subset of $\mathcal{A}$. More specifically, we have to consider only those for which $x_A = k-1, k, k+1, \ldots, n$ (see Figure 4). In other words, an allowable prefix has $k-1$ U steps and $0, 1, 2, \ldots$ or $n-k+1$ D steps. If $b_{i,j}$ denotes the numbers of Dyck prefixes

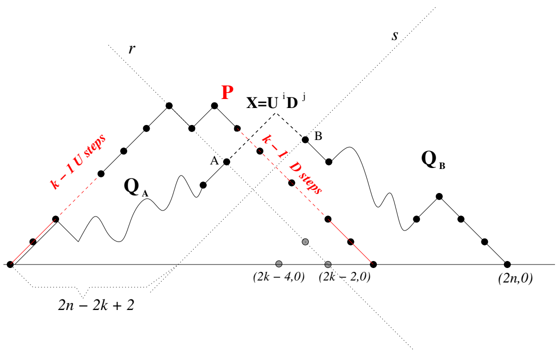

Figure 4: Avoiding $U^{k-1}DUD^{k-1}$, with $x_A \leq x_B$

with i U steps and j D steps ($i \geq j$), then the contribution to $d_n(P)$ in this case is

$$d_n^{(1)}(P) = \left(\sum_{j=0}^{n-k+1} b_{k-2,j} \right)^2 .$$

The coefficients $b_{i,j}$ are the well-known *ballot numbers* (sequence A009766 in [14]), whose first values are reported in Table 1.

b) If $x_A > x_B$, then it is easy to see that Q necessarily avoids P, since A clearly occurs after B, and so there are strictly less than $k-1$ D steps from A to $(2n, 0)$. Observe that, in this case, the path Q lies below the profile drawn by the four lines $y = x$, r, s and $y = -x + 2n$. In order to count these paths, referring to Figure 5, just split each of them into a prefix and a suffix of equal length n and call C the point having abscissa n.

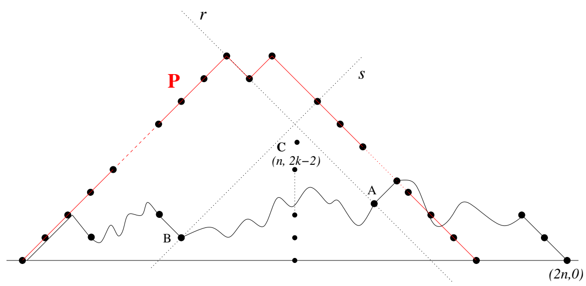

Figure 5: Avoiding $U^{k-1}DUD^{k-1}$, with $x_A > x_B$

Since C must lie under the point where r and s intersect, then its ordinate y_C equals $-n + 2k - 2 - 2t$ with $t \geq 1$ (and also recalling that $y_C = -n + 2k - 2 - 2t \geq 0$). A prefix whose final point is C has $k - j$ U steps and $n - k + j$ D steps, with $j \geq 2$. Since, in this case, a path Q avoiding P is constructed by gluing a prefix and a suffix chosen among $b_{k-j,n-k+j}$ possibilities ($j \geq 2$), we deduce that the contribution to $d_n(P)$ in this case is:

$$d_n^{(2)}(P) = \sum_{j \geq 2} b^2_{k-j,n-k+j}.$$

i \ j	0	1	2	3	4	5	6	7	8	9
0	1									
1	1	1								
2	1	2	2							
3	1	3	5	5						
4	1	4	9	14	14					
5	1	5	14	28	**42**	42				
6	*1*	*6*	*20*	**48**	90	132	132			
7	1	7	**27**	75	165	297	429	429		
8	1	8	35	110	275	572	1001	1430	1430	
9	1	9	44	154	429	1001	2002	3432	4862	4862

Table 1: The sum of the gray entries gives the bold entry in the line below. The sum of the squares of the bold entries gives an appropriate element of Table 2.

Summing up the two contributions we have obtained in a) and b), we get:

$$\begin{aligned} d_n(P) &= d_n^{(1)}(P) + d_n^{(2)}(P) \\ &= \left(\sum_{j=0}^{n-k+1} b_{k-2,j} \right)^2 + \sum_{j \geq 2} b^2_{k-j,n-k+j}. \end{aligned}$$

which is the thesis. ■

Notice that formula (5) reduces to the first sum if $n \geq 2k-3$, since in that case $n-k+j > k-j$, for $j \geq 2$. We then have a single formula including both formulas of the two above lemmas:

$$d_n(P) = \left(\sum_{j=0}^{n-k+1} b_{k-2,j} \right)^2 + \sum_{j \geq 2} b^2_{k-j,n-k+j}, \quad \text{if} \quad n \geq k+1 \quad . \tag{6}$$

Formula (6) can be further simplified by recalling a well known recurrence for ballot numbers, namely that, when $j \leq i+1$,

$$b_{i+1,j} = \sum_{s=0}^{j} b_{i,s}.$$

Therefore, we get the following interesting expression for $d_n(P)$ in terms of sums of squares of ballot numbers along a skew diagonal (see also Tables 1 and 2):

Proposition 3.2 *For $n \geq k+1$,*

$$d_n(P) = \begin{cases} C^2_{k-1} & \textit{if } n \geq 2k-1; \\ \sum_{j \geq 1} b^2_{k-j,n-k+j} & \textit{otherwise.} \end{cases} \tag{7}$$

Therefore we obtain in particular:

$$d_n(UUDUDD) = 4, \text{ when } n \geq 3.$$

k \ n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	...
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...
3	1	1	2	*4*	4	4	4	4	4	4	4	4	4	4	...
4	1	1	2	5	*13*	25	25	25	25	25	25	25	25	25	...
5	1	1	2	5	14	*41*	**106**	196	196	196	196	196	196	196	...
6	1	1	2	5	14	42	*131*	**392**	**980**	1764	1764	1764	1764	1764	...
7	1	1	2	5	14	42	132	*428*	**1380**	**4068**	**9864**	17424	17424	17424	...
8	1	1	2	5	14	42	132	429	*1429*	**4797**	**15489**	**44649**	**105633**	184041	...
9	1	1	2	5	14	42	132	429	1430	*4861*	**16714**	**56749**	**181258**	**511225**	...

Table 2: Number of Dyck paths of semilength n avoiding $U^{k-1}DUD^{k-1}$. Entries in boldface are the nontrivial ones ($k+1 \leq n < 2k-3$).

k \ n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	...
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	...
3	1	1	2	*4*	4	0	0	0	0	0	0	0	0	0	...
4	1	1	2	5	*13*	25	25	0	0	0	0	0	0	0	...
5	1	1	2	5	14	*41*	**106**	196	196	0	0	0	0	0	...
6	1	1	2	5	14	42	*131*	**392**	**980**	1764	1764	0	0	0	...
7	1	1	2	5	14	42	132	*428*	**1380**	**4068**	**9864**	17424	17424	0	...
8	1	1	2	5	14	42	132	429	*1429*	**4797**	**15489**	**44649**	**105633**	184041	...
9	1	1	2	5	14	42	132	429	1430	*4861*	**16714**	**56749**	**181258**	**511225**	...

Table 3: Number of Dyck paths of semilength n avoiding U^kD^k. Entries in boldface are the nontrivial ones ($k+1 \leq n < 2k-3$).

3.3 The pattern U^kD^k

The case $P = U^kD^k$ is very similar to the previous one. We just observe that, when $x_A \leq x_B$, the two points A and B can be connected only using a sequence of D steps followed by a sequence of U steps. This is possible only if $n \leq 2k-2$, which means that r and s do not intersect below the x-axis. Instead, if $n \geq 2k-1$, Q cannot avoid P. Therefore we get (see also Table 3):

Proposition 3.3 *For $n \geq k+1$,*

$$d_n(P) = \begin{cases} 0 & \text{if } n \geq 2k-1; \\ \sum_{j\geq 1} b^2_{k-j,n-k+j} & \text{otherwise.} \end{cases}$$

In particular, we then find:

- $d_n(UUDD) = 0$, when $n \geq 3$;
- $d_n(UUUDDD) = 0$, when $n \geq 5$.

3.4 The pattern $U^{k-1}D^{k-1}UD$

This is by far the most challenging case.

Let Q be a Dyck path of length $2n$ and $P = U^{k-1}D^{k-1}UD$. If Q avoids P, then there are two distinct options: either Q avoids $U^{k-1}D^{k-1}$ or Q contains such a pattern. In the first case, we already know that $d_n(U^{k-1}D^{k-1})$ is eventually equal to zero. So, for the sake of simplicity, we will just find a formula for $d_n(P)$ when n is sufficiently large, i.e. $n \geq 2k-3$. Therefore, for the rest of this section, we will suppose that Q contains $U^{k-1}D^{k-1}$.

The $(k-1)$-th D step of the first occurrence of $U^{k-1}D^{k-1}$ in Q lies on the line having equation $y = -x + 2n$. This is due to the fact that Q has length $2n$ and there cannot be any occurrence of UD after the first occurrence of $U^{k-1}D^{k-1}$. The path Q touches the line of equation $y = -x + 2k - 2$ for the first time with the end point A of its $(k-1)$-th U step. After that, the path Q must reach the starting point B of the $(k-1)$-th D step occurring after A. Finally, a sequence of consecutive D steps terminates Q (see Figure 6). Therefore, Q can be split into three parts: the first part, from the beginning to A, is a Dyck prefix having $k-1$ U steps and ending with a U step; the second part, from A to B, is a path using $n-k+1$ U steps and $k-2$ D steps; and the third part, from B to the end, is a sequence of D steps (whose length depends on the coordinates of A). However, both the first and the second part of Q have to obey some additional constraints.

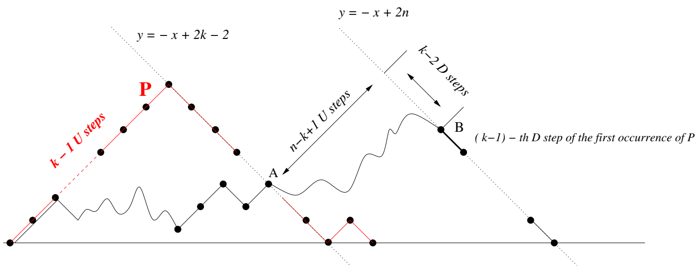

Figure 6: A path Q avoiding $P = U^{k-1}D^{k-1}UD$

The height of the point A (where the first part of Q ends) must allow Q to have at least $k-1$ D steps after A. Thus, the height of A plus the number of U steps from A to B minus the number of D steps from A to B must be greater than or equal to 1 (to ensure that the pattern $U^{k-1}D^{k-1}$ occurs in Q). Hence, denoting with x the maximum number of D steps which can occur before A, either $x = k-2$ or the following equality must be satisfied:

$$(k-1) - x + (n-k+1) - (k-2) = 1.$$

Therefore, $x = \min\{n-k+1, k-2\}$. Observe however that, since we are supposing that $n \geq 2k-3$, we always have $x = k-2$.

Concerning the part of Q between A and B, since we have to use $n-k+1$ U steps and $k-2$ D steps, there are $\binom{n-1}{k-2}$ distinct paths connecting A and B. However, some of them must be discarded, since they fall below the x-axis. In order to count these "bad" paths, we split each of them into two parts. Namely, if A' and B' are the starting and ending points of the first (necessarily D) step below the x-axis, the part going from A to A', and the remaining part (see Fig. 7). It is not too hard to realize that the number of possibilities we have to choose the first part is given by a ballot number (essentially because, reading the path from right to left, we have to choose a Dyck prefix from A' to A), whereas the number of possibilities we have to choose the second part is given by a binomial coefficient (essentially because, after having discarded the step starting at A', we have to choose an unrestricted path from B' to B). After a careful inspection, we thus get to the following result for the total number $d_n(P)$ of Dyck paths avoiding P:

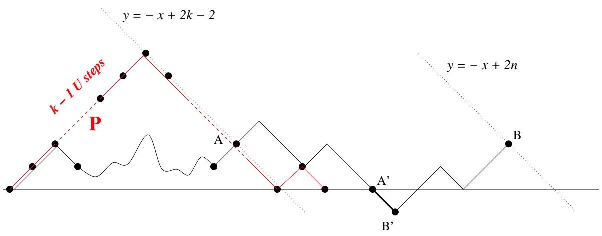

Figure 7: A forbidden subpath from A to B.

$k \backslash n$	0	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1
3	1	1	2	4	6	8	10	12	14	16
4	1	1	2	5	13	28	48	73	103	138
5	1	1	2	5	14	41	110	245	450	739
6	1	1	2	5	14	42	131	397	1069	2427

Table 4: Avoiding $U^{k-1}D^{k-1}UD$

Proposition 3.4 *For $n \geq 2k-3$,*

$$\begin{aligned} d_n(P) = & \binom{n-1}{k-2} C_{k-1} \\ & - \sum_{s=2}^{k-2} b_{k-2,s} \cdot \left(\sum_{i=0}^{s-2} b_{k-3-i,s-2-i} \binom{n-k-s+3+2i}{i} \right). \end{aligned} \tag{8}$$

Formula (8) specializes to the following expressions for low values of k (see also Table 4):

- when $k=3$, $d_n(P) = 2n-2$ for $n \geq 3$;
- when $k=4$, $d_n(P) = \frac{5n^2-15n+6}{2}$ for $n \geq 5$;
- when $k=5$, $d_n(P) = \frac{14n^3-84n^2+124n-84}{6}$ for $n \geq 7$.

4 On the generating function of Dyck paths avoiding one pattern

The goal of this section is to compute the generating functions of Dyck paths avoiding a certain pattern P in a recursive fashion. From these generating function, we get the exact enumeration of these Dyck paths and automatically recover many of the previous results.

Let $\Delta_P(x)$ be the generating function of Dyck paths avoiding P, where x takes into account the length rather than the semilength (thus, $\Delta_P(x)$ is an even power series). We also define intermediate generating functions: let $C_P(x,y)$ be the bivariate generating function of smallest Dyck prefixes containing the pattern P (i.e., such that no proper prefix contains P), where x takes into account the length and y the final height. Note that, for what concerns $C_P(x,y)$, the pattern P can be any Dyck prefix (rather than a Dyck path). Finally, we denote with ϵ the empty path.

Theorem 4.1 *The generating function $C_P(x,y)$ satisfies the following recurrence formulas:*

$$C_\epsilon(x,y) = 1; \tag{9}$$

$$C_{PU}(x,y) = \frac{yC_P(x,y) - xC_P(x,x)}{y-x}xy; \tag{10}$$

$$C_{PD}(x,y) = \frac{xy^{-1}}{1-xy}C_P(x,y) - xy^{-1}C_P(x,0). \tag{11}$$

Moreover, the generating function $\Delta_P(x)$ is given by:

$$\Delta_\epsilon(x) = 0; \tag{12}$$

$$\Delta_{PU}(x) = \Delta_P(x) + C_P(x,x); \tag{13}$$

$$\Delta_{PD}(x) = \Delta_P(x) + C_P(x,0). \tag{14}$$

Proof. Let us start with the generating function $C_P(x,y)$. If $P = \epsilon$, the only smallest prefix containing P is the empty path, with generating function 1. For any fixed path P, let now Q be a smallest Dyck prefix containing $P\,U$. Let Q' be the smallest prefix of Q containing P and let h be the final height of Q'. The path Q is of the form:

$$Q = Q'\,D^i\,U,$$

with $0 \le i \le h$. Its final height is $h - i + 1$. In terms of generating functions, we have:

$$C_{PU}(x,y) = \left(\sum_{h\ge 0}[y^h]C_P(x,y)\right)\sum_{i=0}^{h} x^{i+1}y^{h-i+1},$$

which boils down to (10).

Similarly, let Q be a smallest Dyck prefix containing $P\,D$ and let Q' be its smallest prefix containing P and let h be the final height of Q'. We have:

$$Q = Q'\,U^i\,D.$$

If $h > 0$, we have no restriction on i; however, if $h = 0$, we must have $i > 0$. This yields

$$C_{PD}(x,y) = C_P(x,y)\sum_{i\ge 0}(xy)^i xy^{-1} - C_P(x,0)xy^{-1},$$

which is equivalent to (11).

Consider now the generating function $\Delta_P(x)$. As no path avoids the empty pattern, we have $\Delta_\epsilon(x) = 0$. Let Q be a Dyck path avoiding $P\,U$. We distinguish two cases: either it avoids P, in which case it is enumerated by $\Delta_P(x)$, or it contains P. In the latter case, let Q' be the smallest prefix of Q containing P. As Q avoids PU, we must have:

$$Q = Q'\,D^i,$$

which means that Q is enumerated by $C_P(x,x)$.

Finally, let Q be a path avoiding $P\,D$. If Q also avoids P, it is enumerated by $\Delta_P(x)$; if not, let Q' be the smallest prefix containing P. As Q avoids PD, since a Dyck path must end with a D step, we must have $Q = Q'$. This means that Q is enumerated by $C_P(x,0)$. This completes the proof. ∎

Using this theorem, we can recursively compute the generating function $\Delta_P(x)$ for any pattern P. We now show that we can use it to find a formula for the number $d_P(n)$ of Dyck paths avoiding P.

Corollary 4.1 *For any pattern P, the generating function $C_P(x,y)$ is of the form:*

$$C_P(x,y) = \frac{A(x,y)}{(1-x^2)^i(1-xy)^j}, \tag{15}$$

where $A(x,y)$ is a polynomial and i and j are nonnegative integers. Moreover, the generating function $\Delta_P(x)$ has the form:

$$\Delta_P(x) = \frac{B(x)}{(1-x^2)^k}, \tag{16}$$

where $B(x)$ is an even polynomial and k a nonnegative integer.

Proof. It is clear from Theorem 4.1 that the generating functions $C_P(x,y)$ and $\Delta_P(x)$ are rational.

Consider the denominator of $C_P(x,y)$. Using an inductive argument, it is immediate to see, from Equations (10) and (11), that the denominator of the resulting generating functions can contain in general the following factors: $1-x^2$, $1-xy$, $y-x$, y. However, the two factors $y-x$ and y must be cancelled by the numerator, since $C_P(x,y)$ is well defined as a formal power series in x and y, whence we get an expression as in (15).

An analogous argument shows that the denominator of $\Delta_p(x)$ can contain only the factor $1-x^2$, thus obtaining an expression as in (16). Moreover, since $\Delta_P(x)$ counts Dyck paths, it is an even power series, which means that $B(x)$ is an even polynomial. ■

Equation (16) enables us to compute the number $d_P(n)$ for sufficiently large values of n. Writing $B(x) = B_0 + B_1x^2 + \cdots + B_dx^{2d}$,we have:

$$d_P(n) = \sum_{i=0}^{d} B_i \binom{n-i+k-1}{k-1},$$

which shows that $d_P(n)$ is always given by a polynomial function of n for large values of n.

Finally, we note that it is easy to write a computer program that computes the generating function $\Delta_P(x)$ from the formulas of Theorem 4.1. From this generating function, finding the polynomial giving $d_P(n)$ is also automatic, enabling us to automatically recover many of the previous exact enumeration results.

5 On the asymptotics of pattern avoiding Dyck paths

In the present section we describe the asymptotic behavior of the integer sequences counting Dyck paths avoiding a single path. Unlike what happens for permutations, we will be able to prove a sort of "master theorem", meaning that all the sequences which count Dyck paths avoiding a single pattern P have the same asymptotic behavior (with some parameters, such as the leading coefficient, depending on the specific path P). Our result has a slightly more general scope, since it concerns patterns which are unrestricted words on the alphabet $\{U, D\}$ rather than just Dyck words. We also remark that the coefficient α_P appearing in the statement of the next theorem is different from the conjectured value in [1].

Before stating the theorem, we need to give one technical definition.

Given any two unrestricted paths P and Q on $\{U, D\}$ having the same length, we say that Q is *strictly higher* than P whenever P lies below Q and the only contact points are the starting and ending points of the two paths. Note that, if Q is strictly higher than P, then necessarily Q starts with U and P starts with D, and analogously Q ends with D and P ends with U.

Theorem 5.1 *Let P be a word of $\{U, D\}^*$ of the form $P = U^a\, P'\, D^b$, such that either P' is empty or it starts with a D, ends with a U and has c up steps and d down steps (note that P does not have to be a Dyck path). Let $k = c + d - 2$ and let α_P be the number of paths with c up steps and d down steps strictly higher than P'. The number $d_n(P)$ of Dyck paths of length $2n$ avoiding P is asymptotic to:*

$$d_n(P) \sim \frac{C_a\, C_b\, \alpha_P}{k!} n^k.$$

Proof. We establish the result by induction on k. We first show the result for $k = -2$, i.e. $P = U^a\, D^b$. In this case, looking at the case $P = U^k D^k$ from the previous section and using a completely analogous argument, we know that there are finitely many P-avoiding Dyck paths, which means that $d_n(P)$ is asymptotic to zero; since α_P is zero if P' is the empty path, the theorem holds. Moreover, notice that the hypotheses imply that the case $k = -1$ can never happen. Finally, the case $k = 0$ is again similar to the case $U^{k-1}\, D\, U\, D^{k-1}$ treated in the previous section; in this case we have $\alpha_P = 1$, and it is easy to see that $d_n(P) \sim C_a C_b$, as predicted by our formula.

Assume now that $k > 0$. Let P^- be the pattern P with the final U step of P' deleted. The induction hypothesis entails that the number of Dyck paths of semilength n avoiding P^- is $o(n^k)$; therefore, we can restrict ourselves to the paths that avoid P but not P^-.

Let Q be a Dyck path avoiding P but containing P^-. Let S be the smallest prefix of Q containing the first a up steps and let T be the smallest suffix containing the last b down steps. We have already shown in the previous section that there are C_a possible choices for S and C_b possible choices for T. Moreover, since S and T have length at most $2a$ and $2b$, respectively, they do not intersect for sufficiently large n.

Now, write $P' = p'_1 \cdots p'_{k+2}$ (i.e. p'_i is the i-th step of P') and let Q_i be, for $0 \le i \le k+1$, the smallest prefix of Q containing the pattern $U^a\, p'_1 \cdots p'_i$. By definition of S, we have:

$$Q_0 = S. \tag{17}$$

Moreover, for $1 \le i \le k+1$, the word Q_i is equal to Q_{i-1} extended to the next occurrence of the letter p'_i. In other words, there exists a nonnegative integer ℓ_i such that:

$$Q_i = Q_{i-1} \begin{cases} D^{\ell_i} U & \text{if } p'_i = U; \\ U^{\ell_i} D & \text{if } p'_i = D. \end{cases} \tag{18}$$

Finally, since Q does not contain P and since $p'_{k+2} = U$, there cannot be a U step between the end of the prefix Q_{k+1} and the start of the suffix T. Therefore, we have:

$$Q = Q_{k+1} D^{\ell_{k+2}}\, T. \tag{19}$$

Let $m_1, \ldots, m_c$ and $n_1, \ldots, n_d$ be the sublists of $\ell_1, \ldots, \ell_{k+2}$ corresponding to the indices i such that $p'_i = U$ and $p'_i = D$, respectively. Let $\tilde{m}_1, \ldots, \tilde{m}_c$ and $\tilde{n}_1, \ldots, \tilde{n}_d$ be their partial sums:

$$\tilde{m}_i = m_1 + \cdots + m_i, \qquad\qquad \tilde{n}_i = n_1 + \cdots + n_i.$$

Equations (17), (18) and (19) show that Q is entirely determined by the words S and T and the values $m_1, \ldots, m_c$ and $n_1, \ldots, n_d$ (or, equivalently, by their partial sums). Moreover, since Q has n up steps and n down steps, the complete sums $\tilde{m}_c$ and $\tilde{n}_d$ are fixed; specifically:

$$\tilde{m}_c = n - |S|_D - b - d; \qquad\qquad \tilde{n}_d = n - a - |T|_U - c + 1.$$

Note that both these values are asymptotic to n.

Let $\mathcal{X}$ be the set $\{\tilde{m}_1, \ldots, \tilde{m}_{c-1}, \tilde{n}_1, \ldots, \tilde{n}_{d-1}\}$. As we are dealing with asymptotics, we assume that no two elements of $\mathcal{X}$ are apart by less than $a+b+c+d$ (i.e. the length of P), since the number of possible sets $\mathcal{X}$ where this is not the case is of an order less than n^k. Let R be the path $U\, r_1 \cdots r_k\, D$, with $r_i = D$ if the i-th element of $\mathcal{X}$ in ascending order is one of the $\tilde{m}_j$'s and U if it is one of the $\tilde{n}_j$'s. The path Q is thus determined by the paths S and T, the set $\mathcal{X}$, and the path R.

We now prove that Q is a Dyck path if and only if the path R is strictly higher than P'. First, we note that Q is a Dyck path if and only if the prefixes $Q_1, \ldots, Q_{k+1}$ end at a positive height (the local minima of Q not belonging to S occur only at these points). Let $1 \leq i \leq k+1$ and assume that the first i steps of P' consist of x up steps and y down steps. From the equations (17) and (18), we get the number of U and D steps in Q_i:

$$|Q_i|_U = a + x + \tilde{n}_y; \qquad\qquad |Q_i|_D = |S|_D + y + \tilde{m}_x.$$

Since $\tilde{m}_x$ and $\tilde{n}_y$ are more than the length of P apart, the condition that Q_i ends at a positive height is equivalent to $\tilde{n}_y > \tilde{m}_x$. This, in turn, is equivalent to say that there are at least x up steps in the word $r_1 \cdots r_{i-1}$, that is at least $x+1$ up steps in the first i letters of R. Therefore, Q is a Dyck path if and only if R is strictly higher than P'.

Summing up, we can thus assert that the path Q is determined by:

- the paths S and T (there are C_a and C_b possible choices, respectively);
- the set $\mathcal{X}$ (the number of choices is asymptotic to $\binom{n}{k} \sim \frac{n^k}{k!}$);
- the path R (there are α_P choices, since a path strictly higher than P' must start with a U and end with a D).

This gives the announced estimate. ■

6 Conclusions and further works

In the present paper we have initiated the study of a new poset, the Dyck pattern poset, whose introduction is motivated by trying to find a correct analog of the permutation pattern poset in the case of Dyck paths. However we have only scratched the surface of this subject, and many things still remain to be done. For instance, the poset structure of the Dyck pattern poset certainly needs to be better understood. Our final result is a first step towards this direction.

Theorem 6.1 *The Dyck pattern poset is a* partial well order*, i.e. it contains neither an infinite properly decreasing sequence nor an infinite antichain.*

Proof. This is a consequence of a theorem by Higman [9]. In fact, such a theorem implies the following statement (see for instance [4]): the subword order over a finitely generated free monoid is a partial well order. Since the Dyck pattern containment order is a subposet of the subword order on a two-letter alphabet, the conclusion immediately follows. ■

We close our work with a (not at all exhaustive) list of open problems, concerning both enumerative and order-theoretical issues. Part of the items of the list have been suggested by referees, who are warmly thanked for this.

- What about the enumeration of Dyck paths avoiding two or more patterns?

- Similarly to what has been done for permutations, we may declare two Dyck paths P and Q to be *Wilf-equivalent* whenever $d_n(P) = d_n(Q)$, for all n. The only case of a trivial Wilf-equivalence comes from taking the mirror image of a Dyck path. Are there any nontrivial examples of Wilf-equivalence relations?

- Dyck paths are members of the Catalan family, and so there is plenty of bijections between them and other important combinatorial structures. Can we transport the pattern containment order on Dyck paths along some of these bijections in order to obtain interesting order structures on different combinatorial objects? For instance, in the already mentioned paper [12], Rowland defines (consecutive) patterns on binary trees, and he explores some similarities between his notion and the notion of *factor* (i.e. consecutive subword) on Dyck words. Since our patterns are not consecutive, there seems to be no direct connection with the work of Rowland. It would probably be more interesting to investigate possible analogies with the notion of (non consecutive) pattern on binary trees provided in [5].

- What is the Möbius function of the Dyck pattern poset (from the bottom element to a given path? Of a generic interval?)?

- How many (saturated) chains are there up to a given path? Or in a general interval? Can we say anything more precise on the order structure of intervals (for instance, is it possible to determine when they are lattices?)?

- In the same spirit of the present paper, one can address analogous problems on the posets defined by different types of paths. What immediately comes into mind is to investigate properties of what can be called the Motzkin pattern poset and the Schröder pattern poset. A first look at these two cases shows that they actually constitute only one more case, in the sense that the Motzkin pattern poset and the Schröder pattern poset are isomorphic.

References

[1] A. Bernini, L. Ferrari, R. Pinzani, J. West, *Pattern avoiding Dyck paths*, Discrete Math. Theoret. Comput. Sci. Proc., AS (2013) 683–694.

[2] A. Björner, *The Möbius function of subword order*, Invariant theory and tableaux (Minneapolis, MN, 1988), 118124, IMA Vol. Math. Appl., 19, Springer, New York, 1990.

[3] A. Burstein, *Enumeration of words with forbidden patterns*, PhD thesis, University of Pennsylvania, 1998.

[4] C. Choffrut, J. Karhumäki, *Combinatorics of words*, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Springer, 1997, pp. 329–438.

[5] M. Dairyko, L. Pudwell, S. Tyner, C. Wynn, *Non-contiguous pattern avoidance in binary trees*, Electron. J. Combin., 19(3) (2012) #P22.

[6] L. Ferrari, R. Pinzani, *Lattices of lattice paths*, J. Statist. Plann. Inference, 135 (2005) 77–92.

[7] S. Giraudo, *Balanced binary trees in the Tamari lattice*, Discrete Math. Theoret. Comput. Sci. Proc., AN (2010) 593–604.

[8] A. Goyt, *Avoidance of partitions of a three element set*, Adv. Appl. Math., 41 (2008) 95–114.

[9] G. Higman, *Ordering by divisibility in abstract algebras*, Proc. London Math. Soc., 3 (1952) 326–336.

[10] M. Klazar, *On abab-free and abba-free sets partitions*, European J. Combin., 17 (1996) 53–68.

[11] D. Knuth, *The art of computer programming*, Vol. 1, Addison Wesley, Boston, 1968.

[12] E. Rowland, *Pattern avoidance in binary trees*, J. Combin. Theory Ser. A, 117 (2010) 741–758.

[13] B. E. Sagan, *Pattern avoidance in set partitions*, Ars Combin., 94 (2010) 79–96.

[14] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, electronically available at `oeis.org`.

[15] D. A. Spielman, M. Bóna, *An infinite antichain of permutations*, Electron. J. Combin., 7 (2000) #N2 (4 pp.).