Séminaire Lotharingien de Combinatoire XX (2020)
Article #YY, 12 pp.

Proceedings of the 32nd Conference on Formal Power Series and Algebraic Combinatorics (Ramat Gan)

Generating functions for lattice paths with several forbidden patterns

Andrei Asinowski* and Cyril Banderier† and Valerie Roitner‡

Abstract. This work studies directed lattice paths on $\mathbb{Z}^2$, constrained to avoid a set of given patterns. We give the corresponding generating functions, for walks, meanders (walks additionally constrained to be above the x-axis), and excursions (meanders constrained to end on the x-axis). Our method relies on a vectorial generalization of the classical kernel method, and on a matricial generalization of the autocorrelation polynomial. We apply our approach on more than 512 different models, thus unifying/extending many previous works.

Résumé. Nous étudions les chemins dirigés sur $\mathbb{Z}^2$, contraints à éviter un ensemble de motifs. Nous explicitons les séries génératrices des marches, des méandres (les marches contraintes à rester positives), et des excursions (les méandres contraints à finir à altitude 0). Notre approche utilise une généralisation vectorielle de la méthode du noyau et une généralisation matricielle du polynôme d'autocorrélation. Nous appliquons notre méthode sur plus de 512 modèles différents, unifiant/généralisant ainsi de nombreuses études précédentes.

תקציר. אנחנו אוהבים קומבינטוריקה ואנחנו חוקרים תבניות אסורות במסלולי סריג.

Keywords: *Lattice path, pushdown automaton, forbidden pattern, vectorial kernel method*

1 Definitions and notations

Let $\mathcal{S}$, the *set of steps*, be some finite subset of $\mathbb{Z}$ that contains at least one negative and at least one positive number. A *lattice path with steps from* $\mathcal{S}$ is a finite word $w = (s_1, s_2, \ldots, s_n)$ in which all letters belong to $\mathcal{S}$, visualized as a directed polygonal line in the plane, which starts at the origin and is formed by successive appending of vectors $(1, s_1), (1, s_2), \ldots, (1, s_n)$. The n letters that form the path $w = (s_1, s_2, \ldots, s_n)$ are referred to as its *steps*. The *length* of w, denoted by $|w|$, is the number of steps in w. The *final altitude* of w, denoted by $h(w)$, is the sum of all steps in w, that is $s_1 + s_2 + \ldots + s_n$. Visually, $(|w|, h(w))$ is the point where w terminates. A *pattern* is a fixed word with letters form $\mathcal{S}$. In this article and the companion article [2], we consider lattice paths in which we forbid or mark occurrences of *several patterns* seen as factors in the word representation of the paths (the case of *one pattern* was studied in [1]).

*https://me.aau.at/~anasinowski https://orcid.org/0000-0002-0689-0775
†https://lipn.fr/~banderier https://orcid.org/0000-0003-0755-3022
‡https://dmg.tuwien.ac.at/fb5 https://orcid.org/0000-0002-2621-431X

Andrei Asinowski is supported by the Austrian Science Fund (FWF), in the framework of the project *Analytic Combinatorics: Digits, Automata and Trees* (P 28466-N35).

Under this setting, it is usual to consider two restrictions: that the whole path is (weakly) above the x-axis, and that it has final altitude 0 (equivalently, terminating at the x-axis). Consequently, one considers four classes of lattice paths:

1. A *walk* is any path as described above.
2. A *bridge* is a path that terminates at the x-axis.
3. A *meander* is a path that stays (weakly) above the x-axis.
4. An *excursion* is a path that stays (weakly) above the x-axis and terminates at the x-axis.

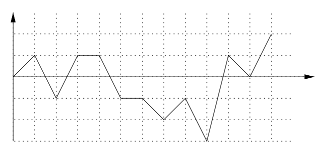
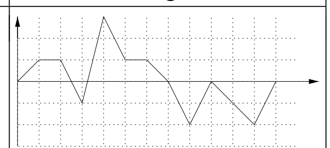
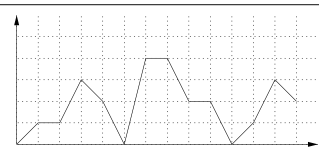
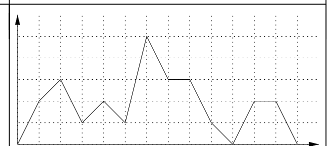
	ending anywhere	ending at 0
on $\mathbb{Z}$	 walks $W(t,u) = \dfrac{\Delta(t,u)}{K(t,u)}$	 bridges $B(t) = -\sum_{i=1}^{e} \dfrac{u_i'}{u_i} \dfrac{\Delta(t,u_i)}{K_t(t,u_i)}$
on $\mathbb{N}$	 meanders $M(t,u) = \dfrac{\Delta(t,u)}{u^e K(t,u)} \prod_{i=1}^{e} (u - u_i(t))$	 excursions $E(t) = \dfrac{(-1)^{e+1}}{t} \prod_{i=1}^{e} u_i(t)$

Table 1: For the four types of paths and for any set of steps encoded by $S(u)$, we give the corresponding generating function of such lattice paths avoiding a set of patterns $p_1, \ldots, p_m$. The formulas involve the e small roots u_i (i.e. $u_i(t) \sim 0$ for $t \sim 0$) of the kernel $K(t,u) := (1 - tS(u))\Delta + \Delta'$, where Δ and Δ' are determinants related to the mutual correlation matrix of the patterns. (See Theorems 1, 2, and 3)

For each of these classes (when no pattern is forbidden), Banderier and Flajolet [5] gave general expressions for the corresponding generating functions and the asymptotics of their coefficients. This study was generalized by Asinowski et al. [1] to the case where the paths are constrained to avoid one single pattern. In this article, we further generalize this last work to the case where the paths are constrained to avoid several patterns. While it was expected that the generating functions would be algebraic, it is a pleasant surprise that they have a nice closed-form, involving some combinatorial determinants, and directly generalizing/unifying all our previous results (and the easier cases of the rational world: patterns in regular expressions, etc.).

Throughout our article, in the generating functions, the variable t corresponds to the length of a path, and the variable u to its final altitude. $S(u)$ is the *step polynomial* associated to the set of steps $\mathcal{S}$, defined by $S(u) = \sum_{s \in \mathcal{S}} u^s$. We also assume that none of the forbidden patterns p_i is a substring of another pattern p_j. (There is no loss of generality in this assumption, since otherwise we can restrict the set of patterns to the set of its minimal elements.)

2 Generating function for walks with forbidden patterns

Theorem 1. *The generating function of walks with steps encoded by $S(u)$ and avoiding the patterns $p_1, \dots, p_m$ is given by*

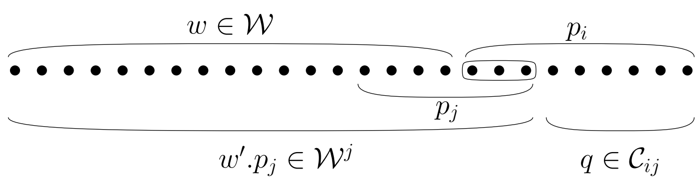
$$W(t,u) = \frac{\Delta(t,u)}{(1 - tS(u))\Delta(t,u) + \sum_{k=1}^{m} t^{|p_i|} u^{\mathrm{alt}(p_i)} \Delta_i(t,u)}, \tag{2.1}$$

where Δ is the determinant of the **mutual correlation matrix** *C defined in (2.4), and Δ_i is defined in (2.5).*

Proof. Let $\mathcal{W}$ be the set of walks avoiding the patterns $p_1, \dots, p_m$, and let $W(u,t)$ be the generating function of $\mathcal{W}$. Further, let $W^{(i)}(t,u)$ be the generating function of all walks that have one occurrence of p_i at the end, but no occurrence of p_i before, as well as no occurrence of any of the other patterns p_j. If we append one step from $\mathcal{S}$ to a walk from $\mathcal{W}$, we either obtain another walk in $\mathcal{W}$, or a walk with one single occurrence of a (uniquely determined) pattern p_i at the end, thus counted by $W^{(i)}$. In terms of generating functions, this means

$$1 + WtS(u) = W + \sum_{j=1}^{m} W^{(j)}. \tag{2.2}$$

Now we take a walk $w \in \mathcal{W}$ and append a pattern p_i to it. Let q be the maximal (possibly empty) suffix of $w.p_i$ such that $w.p_i = w'.p_j.q$, where $w' \in W$ and p_j is one of the forbidden patterns (possibly $p_j = p_i$). Then q is the complement (in p_i) of an *overlap* of p_j and p_i, as shown:



For each pair p_i, p_j, let $\mathcal{C}_{ij}$ be the set[1] of such words q. Associated to these sets $\mathcal{C}_{ij}$, we define the *mutual correlation polynomials*

$$C_{ij}(t,u) := \textstyle\sum_{q \in \mathcal{C}_{ij}} t^{|q|} u^{\mathsf{alt}(q)}.$$

Note that $C_{ii}(t,u)$ is the classical *autocorrelation polynomial* of p_i, as introduced in the case of one single pattern by Schützenberger [17] for prefix codes and by Guibas and Odlyzko [13] in the context of text searching and string overlaps. Equipped with these notations, we have

$$\begin{cases} Wt^{|p_1|} u^{\mathsf{alt}(p_1)} = \sum_{j=1}^{m} W^{(j)} C_{1j}(t,u) \\ \qquad \vdots \\ Wt^{|p_m|} u^{\mathsf{alt}(p_m)} = \sum_{j=1}^{m} W^{(j)} C_{mj}(t,u). \end{cases} \tag{2.3}$$

[1]For example, for the patterns $p_1 = aabb$ and $p_2 = bba$, we have $\mathcal{C}_{12} = \{abb\}$ and $\mathcal{C}_{21} = \{a, ba\}$.

A key role is thus played by what we call the (combinatorial) *mutual correlation matrix* C,

$$C := \begin{pmatrix} C_{11} & \dots & C_{1m} \\ \vdots & \vdots & \vdots \\ C_{m1} & \dots & C_{mm} \end{pmatrix}. \tag{2.4}$$

By Equation (2.2) we have $W = \frac{1-\sum W^{(i)}}{1-tS}$ where the sum $\sum W^{(i)}$ is determined from (2.3) by Cramer's rule[2]: it is equal to $\frac{W}{\Delta}\left(\sum_{k=1}^{m} t^{|p_i|} u^{\mathrm{alt}(p_i)} \Delta_i\right)$, where $\Delta(t,u) := \det(C)$ and

$$\Delta_i(t,u) := \det(C \text{ with all the polynomials } \{C_{ij}\}_{j=1,\dots,m} \text{ replaced by } 1). \tag{2.5}$$

Putting everything together, we obtain the formula given in the theorem. □

Remark. These walks with forbidden patterns can also be encoded by a Markov chain or, equivalently, by a finite automaton. To obtain the generating function with those approaches would typically require the inversion of a $\ell \times \ell$ matrix with symbolic coefficients, which is costly in time and in memory ($\ell := \sum_{i=1}^{m} |p_i|$ is the sum of the lengths of the m forbidden patterns). It is nice that the formula based on the mutual correlation sets is algorithmically much more efficient, and directly gives the generating function, avoiding those larger costs. Note that this formula can also be established via Goulden and Jackson's *cluster method*, e.g. used in [1, 16].

In addition to the mutual correlation matrix, an object which plays a fundamental role for any lattice path enumeration is the denominator of W, which we call the *kernel* of the model. It is

$$K(t,u) := \mathrm{denom}(W(t,u)) = (1 - tS(u))\Delta(t,u) + \sum_{k=1}^{m} t^{|p_i|} u^{\mathrm{alt}(p_i)} \Delta_i(t,u). \tag{2.6}$$

It has e distinct simple roots $u_1, \dots, u_e$ satisfying $u_i(z) \sim 0$ for $z \sim 0$, which we call the *small roots* of the kernel. Similarly to classical directed lattice paths, we shall see that the generating functions of our pattern constrained lattice paths are algebraic and expressible in terms of these small roots. Here is what it gives for bridges:

Theorem 2. *The generating function for bridges avoiding the patterns $p_1, \dots, p_m$ is*

$$B(t) = -\sum_{i=1}^{e} \frac{u_i'}{u_i} \frac{\Delta(t,u_i)}{K_t(t,u_i)} \tag{2.7}$$

Proof. The generating function for bridges is obtained from $W(t,u)$ via a residue computation

$$B(t) = [u^0]W(t,u) = \frac{1}{2\pi i} \int_{|u|=\varepsilon} \frac{W(t,u)}{u}\, du = \sum_{i}^{e} \mathrm{Res}_{u=u_i(t)} \frac{W(t,u)}{u}.$$

The residues inside the small circle $|u| = \varepsilon$ are exactly the e small roots $u_i(t)$ of $K(t,u)$. This leads to the theorem. □

[2] We use the fact: If $Cx = d$ and $C^\top y = \vec{1}$ (where $\vec{1}$ is the all-1's column vector), then $x^\top \vec{1} = d^\top y$, and we apply Cramer's rule on the latter system.

3 Generating function of meanders with forbidden patterns

Theorem 3. *The generating function for meanders avoiding the patterns $p_1, \ldots, p_m$ is*

$$M(t,u) = \frac{G(t,u)}{u^e K(t,u)} \prod_{i=1}^{e} (u - u_i(t)), \tag{3.1}$$

where $u_1(t), \ldots, u_e(t)$ are all the small roots of the kernel $K(t,u)$ (defined in Formula (2.6)), and $G(t,u)$ is some formal power series in t and polynomial in u (defined in (3.3)).

Proof (sketch). Lattice paths with forbidden patterns can be encoded by an automaton: its states $X_1, X_2, \ldots$ are labelled by prefixes of the patterns such that a walk w is in a state labelled σ if and only if σ is the longest label which is a suffix of w. (See the illustration in Section 3.1). We denote the transition matrix of this automaton by A. Moreover we denote by M_i the generating function for meanders that terminate in state X_i, and let $\vec{\mathbf{M}} = (M_1, M_2, \ldots)$. Then we have

$$\vec{\mathbf{M}} = (1, 0, \ldots) + t\vec{\mathbf{M}}A - \{u^{<0}\}(t\vec{\mathbf{M}}A), \tag{3.2}$$

where $\{u^{<0}\}(t\vec{\mathbf{M}}A)$ is the generating function for paths obtained from a meander by adding a step that makes them cross the x-axis. Such a matricial functional equation can be solved by the *vectorial kernel method* developed in [1]. First, Equation (3.2) is conveniently rewritten as $\vec{\mathbf{M}}(I - tA) = \vec{\mathbf{F}}$, and multiplying by $\vec{\mathbf{v}} := \mathrm{adj}(I - tA)\vec{\mathbf{1}}$, one gets $\vec{\mathbf{M}} \det(I - tA) = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}$. Here, it is legitimate to substitute u by any small root u_i of $\det(I - tA) = K(t,u)$, the kernel from the previous section[3]. The u_i's are therefore also roots of the polynomial $\Phi(t,u) := u^e \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}$; this yields the following factorisation

$$\Phi(t,u) = G(t,u) \prod_{i=1}^{e} (u - u_i), \tag{3.3}$$

and leads to Formula (3.1) for $M(t,u)$. The factor $G(t,u)$ has some case dependent combinatorial interpretations, on which we comment later. □

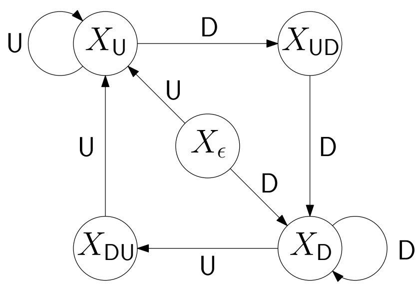
If $\Phi(t,u)$ is a monic polynomial of degree e, then we have the complete factorisation $\Phi(t,u) = \prod_{i=1}^{e}(u - u_i)$ and thus $G(t,u) = 1$: this yields an explicit formula for $M(t,u)$ in terms of $K(t,u)$ and its small roots. It is shown in [1, Section 5] that this happens for some natural cases of a single forbidden pattern.

When one has several patterns, it is generically not the case that $G(t,u) = 1$. In the next sections, we show how, in many cases, one can determine a non-trivial "extra factor" $G(t,u)$. We also show how to obtain the formula for $E(t)$ without computing first $M(t,u)$.

[3]Notice that the equation for *walks* similar to (3.2) reads $\vec{\mathbf{W}} = (1, 0, \ldots) + t\vec{\mathbf{W}}A$. We can solve it and compare $W(t,u) = \vec{\mathbf{W}}\vec{\mathbf{1}} = (1, 0, \ldots)\mathrm{adj}(I - tA)\vec{\mathbf{1}}/\det(I - tA)$ with (2.1). In both expressions, the denominator is a polynomial with constant term 1. Hence $\det(I - tA) = K(t,u)$ and $(1, 0, \ldots)\mathrm{adj}(I - tA)\vec{\mathbf{1}} = \Delta(t,u)$.

3.1 Dyck paths avoiding UDU and DUD

We illustrate the procedure outlined above by the example of Dyck paths avoiding UDU and DUD. These walks are encoded by the following automaton:



Corresponding adjacency matrix
(the states are ordered $X_\epsilon, X_U, X_{UD}, X_D, X_{DU}$):

$$A = \begin{pmatrix} 0 & u & 0 & u^{-1} & 0 \\ 0 & u & u^{-1} & 0 & 0 \\ 0 & 0 & 0 & u^{-1} & 0 \\ 0 & 0 & 0 & u^{-1} & u \\ 0 & u & 0 & 0 & 0 \end{pmatrix}.$$

The kernel $K(t,u) = -u^{-1}(tu^2 - (1+t^2-t^4)u + t)$ can be calculated directly as $\det(I - tA)$, but also by Theorem 2.2 with the mutual correlation matrix $\begin{pmatrix} 1+t^2 & tu \\ t/u & 1+t^2 \end{pmatrix}$. $K(t,u)$ has a unique small root, $u_1(t) = \left(1 + t^2 - t^4 - \sqrt{(1+t+t^2)(1+t-t^2)(1-t+t^2)(1-t-t^2)}\right)/(2t)$.

The functional equation (3.2) has the form

$$(M_1, M_2, M_3, M_4, M_5)(I - tA) = (1,0,0,0,0) - \left(\{u^{<0}\}t(M_1, M_2, M_3, M_4, M_5)A\right).$$

It is easy to see that $\left(\{u^{<0}\}t(M_1, M_2, M_3, M_4, M_5)A\right) = \frac{t}{u}(0,0,0,E(t),0)$ because a path $w.a$ (where w is a meander and a a step) goes below the x-axis if and only if w is an excursion and a is a D-step, and upon making this step the path enters the 4th state X_D. Therefore, the needed components of $\vec{\mathbf{v}} = \text{adj}(I - tA)\vec{\mathbf{1}}$ are $\vec{\mathbf{v}}_1 = 1 + t^2 + t^4$ and $\vec{\mathbf{v}}_4 = 1 + t^3 u$. Thus, the equation $\Phi(t,u) = 0$ has the form

$$(1 + t^2 + t^4)u - tE(t)(1 + t^3 u) = 0. \tag{3.4}$$

Solving it for $E(t)$ and keeping in mind that $u = u_1(t)$ is a root of $\Phi(t,u)$, we obtain

$$E(t) = \frac{u_1(t)(1+t^2+t^4)}{t(1+t^3 u_1(t))} = \frac{1 + t^2 + t^4 - \sqrt{(1+t+t^2)(1+t-t^2)(1-t+t^2)(1-t-t^2)}}{2t^2}.$$

Finally, $G(t,u)$ is simply the leading coefficient of $\Phi(t,u)$ (as polynomial in u), and we can now find it from (3.4). Putting everything together we obtain by Theorem 3 the bivariate generating function for meanders, which in its turn gives the univariate generating function

$$M(t) = -\left(\frac{1-t^3}{2t} - \frac{(1+t)\sqrt{(1+t+t^2)(1+t-t^2)(1-t+t^2)(1-t-t^2)}}{2t(1-t-t^2)}\right).$$

The enumerating sequence for meanders is the sequence `A329703` from the On-Line Encyclopedia of Integer Sequences, and the one for excursions (counted by semilength) is `A004148`, which also counts some other constrained paths (like peakless Motzkin paths – see below), but also some classes of RNA structures, ordered trees, permutations, etc.; see [7, 11, 14].

4 A multi-multivariate generating function for Motzkin paths with any set of forbidden patterns of length two

There is a vast amount of literature on Dyck or Motzkin lattice paths in which some combinations of patterns (like valleys or peaks) are considered. These works often rely on some ad-hoc context-free grammar decompositions; see e.g. [10, 12, 15]. Here, we show how our approach can extend and unify such results by directly finding a generating function with many variables. For example, for Motzkin paths avoiding any combination of forbidden patterns of length 2, one introduces 9 *markers* – auxiliary variables v_p that encode occurrences of all possible patterns p of length 2 (marker v_{UD} for the pattern UD, etc.). This leads to the following theorem.

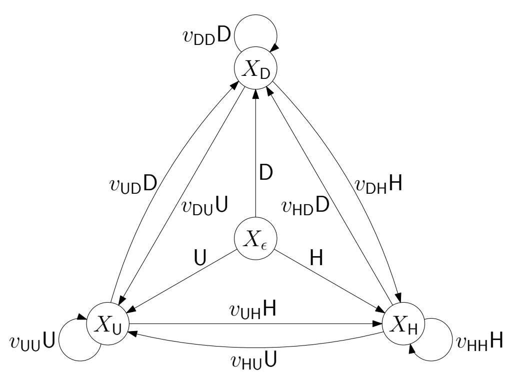
Theorem 4. *The generating function $E(t)$ of Motzkin excursions, where v_p counts the number of occurences of the pattern p, is*

$$\frac{(v_{\mathsf{DD}}-1)-t((v_{\mathsf{DD}}-1)v_{\mathsf{HH}}-(v_{\mathsf{DH}}-1)v_{\mathsf{HD}}-v_{\mathsf{DD}}+v_{\mathsf{DH}})+\Big(1+t(v_{\mathsf{DH}}-v_{\mathsf{HH}})\Big)\frac{u\vec{\mathbf{v}}_1}{t\vec{\mathbf{v}}_4}\Big|_{u=u_1(t)}}{v_{\mathsf{DD}}+t(v_{\mathsf{DH}}v_{\mathsf{HD}}-v_{\mathsf{DD}}v_{\mathsf{HH}})}, \tag{4.1}$$

where $u_1(t)$ is the unique small solution of $\det(I-tA)=0$ for the matrix A defined below, and $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_4$ are the 1st and the 4th components of $\vec{\mathbf{v}} := \operatorname{adj}(I-tA)\vec{\mathbf{1}}$.

Remark. It is striking that all the v_p involving the step U are hidden in $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_4$, and u_1.

Proof. Such paths are encoded by the following automaton and the corresponding adjacency matrix:



$$A = \begin{pmatrix} 0 & u & 1 & u^{-1} \\ 0 & v_{\mathsf{UU}}u & v_{\mathsf{UH}} & v_{\mathsf{UD}}u^{-1} \\ 0 & v_{\mathsf{HU}}u & v_{\mathsf{HH}} & v_{\mathsf{HD}}u^{-1} \\ 0 & v_{\mathsf{DU}}u & v_{\mathsf{DH}} & v_{\mathsf{DD}}u^{-1} \end{pmatrix}$$

From this matrix we compute $K(t,u)=\det(I-tA)$. Generically, $uK(t,u)$ is a polynomial of degree 2 in u, and has one small root, $u_1(t)$. Since a path can cross the x axis only by reading D – hence, entering the 4th state, – only the fourth component of $t\vec{\mathbf{M}}(t)A$ has terms with negative powers of u. Therefore, one has

$$\Phi(t,u) := \vec{\mathbf{v}}_1(t,u) - \vec{\mathbf{v}}_4(t,u)N(t,u) = 0, \tag{4.2}$$

where $\vec{\mathbf{v}}_1(t,u)$ and $\vec{\mathbf{v}}_4(t,u)$ are the first and the fourth components of $\operatorname{adj}(I-tA)\vec{\mathbf{1}}$, and $N(t)$ is the generating function for the terms with negative powers of u in the fourth component of $t\vec{\mathbf{M}}(t)A$.

In order to use (4.2) for computing $E(t)$, we find an equation which relates $N(t,u)$ to $E(t)$. Let $E_{\mathsf{H}}(t)$ and $E_{\mathsf{D}}(t)$ be the generating functions for excursions whose last step is H resp. D. Then we have $N(t,u) = \frac{t}{u}\Big(1 + v_{\mathsf{HD}}E_{\mathsf{H}}(t) + v_{\mathsf{DD}}E_{\mathsf{D}}(t)\Big)$. Further we have $E(t) = 1 + E_{\mathsf{H}}(t) + E_{\mathsf{D}}(t)$ and $E_{\mathsf{H}}(t) = t\Big(1 + v_{\mathsf{HH}}E_{\mathsf{H}}(t) + v_{\mathsf{DH}}E_{\mathsf{D}}(t)\Big)$. This allows us to express N in terms of E. Finally, since u_1 is a root of Φ, the substitution of $u = u_1(t)$ into (4.2) gives the formula for $E(t)$. $\square$

In (4.1), setting $v_p = 1$ allows the pattern p, while setting $v_p = 0$ forbids it. We ran an exhaustive analysis of all the $2^9 = 512$ cases; this leads to the following 75 distinct sequences.

Allowed set of patterns	OEIS[4] entry	GF	Growth rate	Allowed set of patterns	OEIS entry	GF	Growth rate	Allowed set of patterns	OEIS entry	GF	Growth rate
000000000	A019590	pol	0	010011100	A020711†	rat	≈ 1.466	101010101	A329696	alg	2
010001000	A329670	pol	0	010011110	A000930	rat	≈ 1.466	011110101	A329695†	alg	2
010001010	A329677	pol	0	011110001	A020711†	rat	≈ 1.466	101010111	A110199	alg	2
001000000	A130716	pol	0	001110010	A068921	rat	≈ 1.466	011110011	A216604†	alg	2
001000010	A329678	pol	0	010101101	A329687	alg	≈ 1.587	101110011	A329698	alg	2
011001010	A329679	pol	0	011100011	A329688	alg	≈ 1.587	010111011	A023432†	alg	≈ 2.148
010001100	A329680	rat	1	010101011	A329689	alg	≈ 1.618	011111010	A023432†	alg	≈ 2.148
110001001	A135528	rat	1	110011011	A324969	rat	≈ 1.618	101001111	A329699	alg	≈ 2.206
010001110	A011655†	rat	1	011101010	A320690	alg	≈ 1.618	101100111	A329700	alg	≈ 2.206
001000100	A266591	rat	1	011011100	A001611†	rat	≈ 1.618	110011111	A217282†	alg	≈ 2.241
011100001	A329682	rat	1	011011110	A000045†	rat	≈ 1.618	101111011	A217282	alg	≈ 2.241
000010000	A000012	rat	1	110001101	A329691	alg	≈ 1.755	110101111	A329676	alg	≈ 2.247
001100010	A100063	rat	1	011100101	A329692	alg	≈ 1.755	011101111	A329666	alg	≈ 2.247
010011000	A329683	rat	1	101100011	A329693	alg	≈ 1.755	010111111	A023431	alg	≈ 2.315
110011001	A065033	rat	1	010101111	A248100	alg	≈ 1.835	011111011	A023431†	alg	≈ 2.315
010011010	A000027†	rat	1	011101011	A329694	alg	≈ 1.835	101011111	A329701	alg	≈ 2.325
001010000	A329684	rat	1	110001111	A025250†	alg	≈ 1.947	101110111	A329702	alg	≈ 2.325
001010100	A040001	rat	1	011100111	A166289	alg	≈ 1.947	101101111	A007477	alg	≈ 2.383
001010011	A046698†	rat	1	110101011	A329664	alg	2	110111011	A004149†	alg	≈ 2.414
001010110	A008619	rat	1	101000101	A126120†	alg	2	011111110	A004149†	alg	≈ 2.414
011011010	A028310	rat	1	101000111	A208355†	alg	2	101111111	A090344	alg	≈ 2.562
110001011	A000931†	rat	≈ 1.325	110011101	A329695	alg	2	110111111	A004148	alg	≈ 2.618
011001100	A000931†	rat	≈ 1.325	010111101	A216604	alg	2	011111111	A004148†	alg	≈ 2.618
001100110	A000931†	rat	≈ 1.325	010111010	A023426†	alg	2	111101111	A104545	alg	≈ 2.732
010101010	A329686	alg	≈ 1.414	011101110	A329671	alg	2	111111111	A001006	alg	3

Table 2: Motzkin excursions avoiding a set of patterns of length 2. Allowed sets of patterns are here indicated via a binary word of length 9, whose bits correspond to the allowance (or not) of UU, UH, UD, HU, HH, HD, DU, DH, DD (in this order). The column GF indicates whether the generating function is polynomial (pol), rational (rat), or algebraic (alg).

Our exhaustive analysis also shows that the 512 cases lead to 158 distinct sequences for meanders. We give more details at the address `https://lipn.fr/~cb/KernelMethod`. Note the same list for all set of patterns of length 5 would have more entries than the estimated number of atoms in the universe!

[4] All the sequences labelled A329xxx are entries that we added to the On-Line Encyclopedia of Integer Sequences (OEIS), available at https://oeis.org/.The sequences marked by † are in the OEIS, but with a few terms of offset.

Via our approach, it is thus not difficult to get the explicit formulas for $E(t)$ and $M(t)$. These generating functions with all the markers v_p are however quite lengthy. So, we now give these explicit formulas when the set of patterns is included in $\{\mathsf{UU}, \mathsf{HH}, \mathsf{DD}\}$ or in $\{\mathsf{UD}, \mathsf{HH}, \mathsf{DU}\}$.

Hanc marginis exiguitas non caperet!

Forbidden patterns	Generating functions of meanders and excursions	OEIS[5]	Growth rate [6]
UU, HH, DD	$M = -(1+t)\Big((1+t)(1-2t) - \sqrt{1-2t+t^2-4t^3+4t^4}\Big)/\big(2t^2(1-2t)\big)$ $E = (1+t)\Big(1-t^2-2t^3-(1+t)\sqrt{1-2t+t^2-4t^3+4t^4}\Big)/(2t^4)$	A329665 A329671	2
UU, HH	$M = -(1+t)\Big(1-3t^2-t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/\big(2t^2(1-2t-t^2)\big)$ $E = \Big(1-t^2-t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/(2t^3).$	A329667 A329666	$A_{\mathsf{UU,HH}} :=$ $\rho(1-t-2t^2+t^3)$
UU, DD	$M = -\Big((1+t)(1-2t-t^2) - \sqrt{1-2t-t^2-t^4+2t^5+t^6}\Big)/\big(2t^2(1-2t-t^2)\big)$ $E = \Big(1-t-t^2-t^3-\sqrt{1-2t-t^2-t^4+2t^5+t^6}\Big)/(2t^4)$	A308435 A004149†	$1+\sqrt{2}$
HH, DD	$M = -\Big(1-2t-3t^2-t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/\big(2t(1+t)(1-2t-t^2)\big)$ $E = \Big(1-t^2-t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/(2t^3)$	A329669 A329666	$1+\sqrt{2}$ $A_{\mathsf{UU,HH}}$
UU	$M = -(1+t)\Big(1-t-3t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/\big(2t^2(1-2t-2t^2)\big)$ $E = \Big(1-t-t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/(2t^3)$	A329672 A004148†	$(3+\sqrt{5})/2$
HH	$M = -\Big(1-2t-2t^2-\sqrt{1-4t^2-8t^3-4t^4}\Big)/\big(2t(1-2t-2t^2)\big)$ $E = \Big(1-\sqrt{1-4t^2-8t^3-4t^4}\Big)/\big(2t^2(1+t)\big)$	A329673 A104545	$1+\sqrt{3}$
DD	$M = -\Big(1-3t-t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/\big(2t(1-2t-2t^2)\big)$ $E = \Big(1-t-t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/(2t^3)$	A329674 A004148†	$1+\sqrt{3}$ $(3+\sqrt{5})/2$
UD, HH, DU	$M = -(1+t)\Big((1+t)(1-2t) - \sqrt{1-2t+t^2-4t^3+4t^4}\Big)/\big(2t^2(1-2t)\big)$ $E = (1+t)\Big(1-t-\sqrt{1-2t+t^2-4t^3+4t^4}\Big)/(2t^3)$	A329665 A329664	2
UD, HH	$M = -\Big(1-2t-t^2+t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/\big(2t(1-2t-t^2+t^3)\big)$ $E = \Big(1+t^2+t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/\big(2t^2(1+t)\big)$	A329675 A329676	$A_{\mathsf{UU,HH}}$
UD, DU	$M = -\Big((1+t)(1-2t-t^2) - \sqrt{1-2t-t^2-t^4+2t^5+t^6}\Big)/\big(2t^2(1-2t-t^2)\big)$ $E = \Big(1-t-t^2-t^3-\sqrt{1-2t-t^2-t^4+2t^5+t^6}\Big)/(2t^4)$	A308435 A004149†	$1+\sqrt{2}$
HH, DU	$M = -(1+t)\Big(1-t-3t^2+t^4-(1-t)\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/\big(2t^2(1-2t-t^2+t^3)\big)$ $E = \Big(1-t^2-t^3-\sqrt{1-2t^2-6t^3-3t^4+2t^5+t^6}\Big)/(2t^3)$	A329668 A329666	$A_{\mathsf{UU,HH}}$
UD	$M = -\Big(1-3t+t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/\big(2t(1-3t+t^2)\big)$ $E = \Big(1-t+t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/(2t^2)$	A088518* A004148†	$(3+\sqrt{5})/2$
DU	$M = -\Big((1+t)(1-3t+t^2)-(1-t)\sqrt{1-2t-t^2-2t^3+t^4}\Big)/\big(2t^2(1-3t+t^2)\big)$ $E = \Big(1-t-t^2-\sqrt{1-2t-t^2-2t^3+t^4}\Big)/(2t^3)$	A088518* A004148†	$(3+\sqrt{5})/2$
none	$M = -\Big(1-3t-\sqrt{1-2t-3t^2}\Big)/\big(2t(1-3t)\big)$ $E = \Big(1-t-\sqrt{1-2t-3t^2}\Big)/(2t^2).$	A005773† A001006	3

The asymptotics of these sequences is $\frac{C}{\sqrt{\pi}}A^n n^\alpha$, where the constant C and the growth rate A are algebraic numbers which depend on the model. The exponent α is universal: as explained in [1, 4, 5], it is

[5]See footnote 4 on previous page for the † symbol. Also, the sequences marked by * are bisections of A088518 that enumerates "symmetric secondary structures of RNA molecules with n nucleotides"; see [14].

[6]The notation $\rho(P)$ used for some growth rates stands for the largest positive root of the polynomial P.

- $\alpha = -3/2$ for meanders with negative drift (above, when $\{\mathsf{UU}, \mathsf{HH}\}$ or $\{\mathsf{UU}\}$ are forbidden) and for excursions,
- $\alpha = 0$ for meanders with positive drift (above, when $\{\mathsf{HH}, \mathsf{DD}\}$ or $\{\mathsf{DD}\}$ are forbidden),
- $\alpha = -1/2$ for meanders with zero drift.

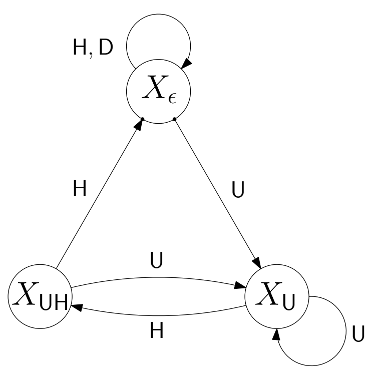
The *drift* is the quantity

$$\delta := \lim_{n\to\infty} \frac{\text{average final altitude of walks on } \mathbb{Z} \text{ of length } n}{n}.$$

Unlike in the article [5], it is no more the case that $\delta = S'(1)$: there is a more subtile interplay between the forbidden patterns and the allowed steps $\mathcal{S}$. In the long version of our article, we comment more on the link between the drift and a weighted variant of the stationary distribution of the automaton A.

Another interesting feature of the above tables is that they suggest there could be natural bijections between different classes of pattern-avoiding Motzkin paths. Some of them are easily seen (for example by mirroring paths), others are less trivial – for example, bijections between UU-avoiding, DU-avoiding, and (with a shift) UD-avoiding Motzkin excursions.

We end this section by mentioning some models involving some patterns of length 3, some appear to be in bijection with some of the previous models. For example, the automaton of Motzkin paths avoiding UD and UHD is



Corresponding adjacency matrix:

$$A = \begin{pmatrix} 1+u^{-1} & u & 0 \\ 0 & u & 1 \\ 1 & u & 0 \end{pmatrix}.$$

Our vectorial kernel method then leads to the fact that the meanders are counted by the "generalized Catalan numbers" A004149, defined by

$$a(n+1) = a(n) + \sum_{k=2}^{n-1} a(k)a(n-1-k), \qquad a(0) = 1.$$

It is noteworthy that it also counts

- Motzkin excursions of length $n-1$ avoiding peaks (UD) and valleys (DU),
- Motzkin excursions of length $n-2$ avoiding UU and UHU.

We leave to the reader the pleasure to find bijective proofs of these facts![7]

[7] See the PhD thesis of the third author for several possible bijections.

5 Beyond rational and algebraic cases

In [3], the authors introduced a notion of patterns in lattice paths similar to the one frequently used for forbidden patterns in permutations: one forbids a sequence of letters which are not necessarily contiguous.

Such a notion is in fact also accessible with our pushdown automaton approach: to forbid a noncontiguous pattern UD is equivalent to avoid the regular expression $\mathsf{U}\mathcal{S}^*\mathsf{D}$. This is achieved by first constructing the automaton for the (contiguous) pattern UD, and then adding a transition labelled by each step from $\mathcal{S}$ to the state X_{U}. This automaton is then easily determinized (to avoid any ambiguity in the transitions), and our vectorial kernel method then handles the additional positivity constraints (meanders, excursions, etc.).

Note that it is a priori not possible to generate (or to forbid) via a finite automaton a pattern which would require the memory of a counter (for example $\mathsf{U}^n\mathsf{H}^n\mathsf{D}^n$). However, we can enumerate such paths by first getting the generating function associated to $\mathsf{U}^i\mathsf{H}^j\mathsf{D}^k$-avoiding paths, and then by taking the diagonal $i = j = k$, at the level of the algebraic generating function (this leads to a D-finite generating function). The same idea allows us to enumerate paths avoiding an infinite set of noncontiguous patterns like e.g. $\{\mathsf{U}^k\mathsf{D}^k\mathsf{UD}\}_{k\in\mathbb{N}}$. We present more examples of this type in the long version of our article.

6 Conclusion and further works

To summarize, in this article we introduced/presented

- the mutual correlation matrix, an extension of the notion of autocorrelation polynomial, which has its own interest and which offers several algorithmic advantages,
- closed-forms for all the main generating function of constrained lattice paths (Sections 2 and 3), generalizing the previous works [1, 5],
- new ways to apply the vectorial kernel method, bypassing the obstacle of the determination of the prefactor G from Equation (3.3),
- new sources for enriched bijections between different types of paths, and the classification of more than 512 models (Section 4).

This will allow us to tackle further questions, like

- faster uniform random generation of constrained paths of length n (Note that the detailed analysis of the Boltzmann method done in [9] is not holding here, because paths with forbidden patterns do not have generically a strongly connected grammar.),
- to mix the analysis of patterns with some additional parameters (like the area below the path [6]) or under further constraints (like to be below a line of rational slope),
- the asymptotics and the Gaussian limit law for the number of occurences of the patterns (another instance of Borges's theorem, see [1, 2, 8]).

References

[1] Andrei Asinowski, Axel Bacher, Cyril Banderier, and Bernhard Gittenberger. Analytic combinatorics of lattice paths with forbidden patterns, the vectorial kernel method, and generating functions for pushdown automata. *Algorithmica*, pages 1–43, 2019.

[2] Andrei Asinowski and Cyril Banderier. Generating functions and multivariate gaussian distribution for lattice paths with marked patterns (extended abstract). In *Proceedings of AofA'20*, pages 24.1–24.15, Klagenfurt, Austria, 2020.

[3] Axel Bacher, Antonio Bernini, Luca Ferrari, Benjamin Gunby, Renzo Pinzani, and Julian West. The Dyck pattern poset. *Discrete Math.*, 321:12–23, 2014.

[4] Cyril Banderier and Michael Drmota. Formulae and asymptotics for coefficients of algebraic functions. *Combinatorics, Probability and Computing*, 24(1):1–53, 2015.

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

[6] Cyril Banderier and Bernhard Gittenberger. Analytic combinatorics of lattice paths: enumeration and asymptotics for the area. *Disc. Math. Theor. Comp. Sci.*, AG:345–355, 2006.

[7] Jean-Luc Baril, Sergey Kirgizov, and Armen Petrossian. Enumeration of Łukasiewicz paths modulo some patterns. *Discrete Math.*, 342(4):997–1005, 2019.

[8] Alberto Bertoni, Christian Choffrut, Massimiliano Goldwurm, and Violetta Lonati. On the number of occurrences of a symbol in words of regular languages. *Theoret. Comput. Sci.*, 302(1-3):431–456, 2003.

[9] Olivier Bodini and Yann Ponty. Multi-dimensional Boltzmann sampling of languages. *Disc. Math. Theor. Comp. Sci.*, AM, January 2010.

[10] Charlotte Brennan and Simon Mavhungu. Visits to level r by Dyck paths. *Fund. Inform.*, 117(1-4):127–145, 2012.

[11] Emeric Deutsch and Louis W. Shapiro. A bijection between ordered trees and 2-Motzkin paths and its many consequences. *Discrete Math.*, 256(3):655–670, 2002.

[12] Sen-Peng Eu, Shu-Chung Liu, and Yeong-Nan Yeh. Dyck paths with peaks avoiding or restricted to a given set. *Stud. Appl. Math.*, 111(4):453–465, 2003.

[13] Leonidas J. Guibas and Andrew M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, 30(2):183–208, 1981.

[14] Emma Y. Jin, Jing Qin, and Christian M. Reidys. Combinatorics of RNA structures with pseudoknots. *Bull. Math. Biol.*, 70(1):45–67, 2008.

[15] Toufik Mansour and Mark Shattuck. Counting humps and peaks in generalized Motzkin paths. *Discrete Appl. Math.*, 161(13-14):2213–2216, 2013.

[16] John Noonan and Doron Zeilberger. The Goulden–Jackson cluster method: Extensions, applications and implementations. *J. Difference Equ. Appl.*, 5(4-5):355–377, 1999.

[17] Marcel-Paul Schützenberger. On the synchronizing properties of certain prefix codes. *Information and Control*, 7:23–36, 1964.