On Lattice Paths with Marked Patterns: Generating Functions and Multivariate Gaussian Distribution

Andrei Asinowski
University of Klagenfurt, Austria
https://me.aau.at/~anasinowski

Cyril Banderier
Université Paris 13 (LIPN, UMR CNRS 7030), France
https://lipn.fr/~banderier

Abstract

In this article, we analyse the joint distribution of some given set of patterns in fundamental combinatorial structures such as words and random walks (directed lattice paths on $\mathbb{Z}^2$). Our method relies on a vectorial generalization of the classical kernel method, and on a matricial generalization of the autocorrelation polynomial (thus extending the univariate case of Guibas and Odlyzko). This gives access to the multivariate generating functions, for walks, meanders (walks constrained to be above the x-axis), and excursions (meanders constrained to end on the x-axis). We then demonstrate the power of our methods by obtaining closed-form expressions for an infinite family of models, in terms of simple combinatorial quantities. Finally, we prove that the joint distribution of the patterns in walks/bridges/excursions/meanders satisfies a multivariate Gaussian limit law.

2012 ACM Subject Classification Mathematics of computing → Generating functions; Mathematics of computing → Distribution functions; Theory of computation → Random walks and Markov chains; Theory of computation → Grammars and context-free languages

Keywords and phrases Lattice path, Dyck path, Motzkin path, generating function, algebraic function, kernel method, context-free grammar, multivariate Gaussian distribution

Digital Object Identifier 10.4230/LIPIcs.AofA.2020.1

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

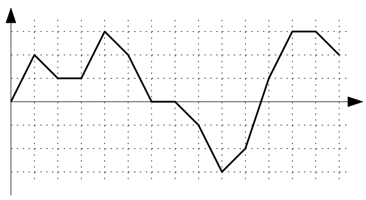
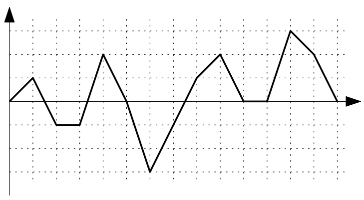
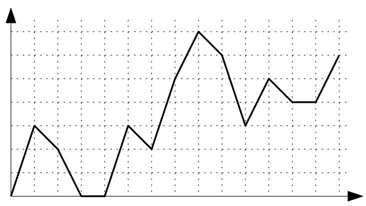
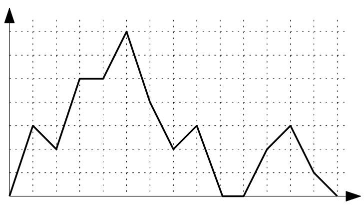
Acknowledgements We thank the referees for their feedback.

1 Definitions and notations for directed lattice paths

Let $\mathcal{S}$, the *set of steps* (or *jumps*), 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 in 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, to be denoted by $\ell(w)$, is the number of steps in w. The *final altitude* of w, to be denoted by $h(w)$, is the sum of all steps in w, that is $s_1 + s_2 + \ldots + s_n$. Visually, $\ell(w)$ and $h(w)$ are the x- and the y-coordinates of the point where w terminates. One considers four classes of paths: a *walk* is any path as described above; a *bridge* is a path that terminates at the x-axis; a *meander* is a path that stays (weakly) above the x-axis; an *excursion* is a path that stays (weakly) above the x-axis and terminates at the x-axis.

31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020).
Editors: Michael Drmota and Clemens Heuberger; Article No. 1; pp. 1:1–1:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Table 1 For the four types of paths (walks, bridges, meanders, excursions) and for any set of steps encoded by $S(u)$, we give the corresponding generating function marking 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 correlation matrix of the patterns. (See Theorems 4, 10, and 12.)

	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 $E(t) = \dfrac{(-1)^{e+1}}{t} \prod_{i=1}^{e} u_i(t)$

For each of these classes (in the simpler case of no pattern constraint), Banderier and Flajolet [6] gave general expressions for the corresponding generating functions and the asymptotics of their coefficients. A unified study of lattice paths with **a single forbidden pattern** was recently started by Asinowski, Bacher, Banderier, and Gittenberger [1]: for any fixed path p (a "pattern") they give the generating function and the asymptotics for paths that avoid p as a consecutive string. Moreover, they initiated the more general analysis of **marking a pattern**: here, one considers a generating function with an extra variable v which encodes the number of occurrences of the pattern p in the path. Setting $v = 0$ gives the generating function for walks that avoid p. In this article, we further generalize this work to the case where **several patterns** are marked. The situation is more challenging: more correlations create more obstacles; however, we shall see that one can still derive closed-form expressions in terms of natural combinatorial quantities!

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* of the set of steps $\mathcal{S}$, defined by

$$S(u) := \sum_{s \in \mathcal{S}} u^s.$$

The set of forbidden/marked patterns will be denoted by $\mathcal{P} = \{p_1, \ldots, p_m\}$.

2 Generating functions for walks and bridges with marked patterns

For the case of a single marked pattern p (see e.g. [1, Thm. 7.1]), the trivariate generating function of walks is

$$W(t,u,v) = \frac{v+(1-v)R}{(1-tS)\big(v+(1-v)R\big)+(1-v)t^{\ell(p)}u^{h(p)}}, \tag{1}$$

where t, u, v are the variables as explained in Section 1; $\ell(p)$ and $h(p)$ are the length and the final altitude of the pattern p; and $R = R(t,u)$ is the autocorrelation polynomial that encodes the overlaps of p with itself – see the definition below. The specialization $v = 0$ gives $W(t,u,0) = R/\big((1-tS)R + t^{\ell(p)}u^{h(p)}\big)$, the generating function of walks that avoid p; and $v = 1$ gives $W(t,u,1) = 1/(1-tS)$ – as expected, since it enumerates all the walks over $\mathcal{S}$.

In this work, we consider the more general case of marking several patterns. To this end, let $\mathcal{S}$ be a set of steps, and let $\mathcal{P} = \{p_1, p_2, \ldots, p_m\}$ be a set of patterns (that is, fixed words over $\mathcal{S}$). In what follows, we assume that $\mathcal{P}$ is a *reduced system*, that is, the words $p_1, p_2, \ldots, p_m$ do not contain each other (where the inclusion is understood as that of strings, for example $\mathsf{ab} \subset \mathsf{abcd}$ and $\mathsf{bc} \subset \mathsf{abcd}$ but $\mathsf{ac} \not\subset \mathsf{abcd}$).

A central role in our approach is played by the notion of *mutual correlation*, a way to formalize how patterns overlap with each other. Given two patterns p_i and p_j, an *overlap* of p_i and p_j is a non-empty string that occurs as a suffix in p_i and as a prefix in p_j. Let $\mathcal{O}_{i,j}$ be the set of all overlaps of p_i and p_j. Further, let $\mathcal{C}_{i,j}$ be the set of words obtained from p_j by erasing all the of overlaps p_i and p_j (as prefixes of p_j). More formally, this leads to the following definition.

▶ **Definition 1** (Mutual correlation polynomials). *The mutual correlation sets are defined as*

$$\mathcal{C}_{i,j} = \{q\colon\ \exists q', q''(q'' \neq \epsilon)\colon\ p_i = q'.q'', p_j = q''.q\}. \tag{2}$$

Accordingly, the ***mutual correlation polynomials*** *are defined as*

$$C_{i,j}(t,u) = \sum_{q\in\mathcal{C}_{i,j}} t^{\ell(q)}u^{h(q)}. \tag{3}$$

In particular, for $i = j$, $C_{i,i}(t,u)$ is the *autocorrelation polynomial* introduced in the case of one single pattern by Schützenberger [38] for prefix codes and by Guibas and Odlyzko [27] in the context of text searching and string overlaps, see also [25, Formula (8.81)] for a first generalization.

▶ **Example 2.** Let $p_1 = \mathsf{aaba}$, $p_2 = \mathsf{abab}$. Then we have

- $\mathcal{O}_{1,1} = \{\mathsf{aaba}, \mathsf{a}\}$, $\mathcal{C}_{1,1} = \{\epsilon, \mathsf{aba}\}$, $C_{1,1} = 1 + t^3u^{2a+b}$;
- $\mathcal{O}_{1,2} = \{\mathsf{aba}, \mathsf{a}\}$, $\mathcal{C}_{1,2} = \{\mathsf{b}, \mathsf{bab}\}$, $C_{1,2} = tu^b + t^3u^{a+2b}$;
- $\mathcal{O}_{2,1} = \mathcal{C}_{2,1} = \varnothing$, $C_{2,1} = 0$;
- $\mathcal{O}_{2,2} = \{\mathsf{abab}, \mathsf{ab}\}$, $\mathcal{C}_{2,2} = \{\epsilon, \mathsf{ab}\}$, $C_{2,2} = 1 + t^2u^{a+b}$.

Let $W = W(t, u, v_1, \ldots, v_m)$ be the generating function for the walks, where each occurrence of the pattern p_i $(i = 1, \ldots, m)$ is marked by the variable v_i. That is, the coefficient of $t^\alpha u^\beta v_1^{\gamma_1} \ldots v_m^{\gamma_m}$ in W is the number of walks of length α and final altitude β that have exactly γ_i occurrences of p_i for $i = 1, \ldots, m$. (Note that occurrences of each pattern are taking self-overlaps into account: thus, for example, the path aaaa contains three occurrences of aa and two occurrences of aaa.)

▶ Remark 3 (The automaton paradigm). Given $\mathcal{S}$ and $\mathcal{P}$, walks with marked patterns can be encoded by a finite automaton $\mathcal{A}$: a walk w is in state Z_α if α is the longest overlap of w with some pattern(s) (if there are no such overlaps, then w is in the initial state Z_ϵ)[1]. This approach leads to the formula

$$W(t,u,v_1,\dots,v_m) = \frac{(1,0,\dots,0)\ \mathrm{adj}(I-tA)\ (1,\dots,1)^\top}{\det(I-tA)}, \tag{4}$$

where $A = A(u,v_1,\dots,v_m)$ is the transition matrix of the automaton $\mathcal{A}$. (NB: the first row/column of A correspond to the initial state Z_ϵ.) In Formula (4), the vector $(1,0,\dots,0)$ encodes the fact that the state Z_ϵ is the single initial state of $\mathcal{A}$, and the vector $(1,\dots,1)^\top$ encodes the fact that all states of $\mathcal{A}$ are final. (See the automaton in Example 9 for an illustration.)

Our first result is another more combinatorial formula for W, bypassing the computational cost inherent to the automaton paradigm approach. This formula can be established via the *cluster method*, as popularized by Goulden and Jackson [24]. It is an instance of what Flajolet called *symbolic inclusion-exclusion* and it was e.g. used in [1,11,32,35,39]. Below, we opt for another proof strategy which emphasizes the role of the mutual correlation polynomials.

▶ **Theorem 4.** *Let $\mathcal{S}$ be a set of steps, and $\mathcal{P} = \{p_1,\dots,p_m\}$ a set of (mutually not included) patterns. The* ***multivariate generating function of walks*** *(where t encodes the length, u the final altitude, and v_i occurrences of the pattern p_i) is given by*

$$W(t,u,v_1,\dots,v_m) = \frac{\Delta}{(1-tS(u))\Delta + \sum_{i=1}^{m} \Delta_i t^{\ell_i} u^{h_i}}, \tag{5}$$

where $\Delta = \Delta(t,u,v_1,\dots,v_m)$ is the determinant of the ***mutual correlation matrix***

$$\begin{pmatrix} v_1+(1-v_1)C_{1,1} & (1-v_2)C_{2,1} & \cdots & (1-v_m)C_{m,1} \\ (1-v_1)C_{1,2} & v_2+(1-v_2)C_{2,2} & \cdots & (1-v_m)C_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ (1-v_1)C_{1,m} & (1-v_2)C_{2,m} & \cdots & v_m+(1-v_m)C_{m,m} \end{pmatrix}, \tag{6}$$

and, for $i = 1,\dots,m$, $\Delta_i = \Delta_i(t,u,v_1,\dots,v_m)$ is the determinant of the matrix obtained from the mutual correlation matrix by replacing its ith row with $(1-v_1,\dots,1-v_m)$, and ℓ_i and h_i are the length and the final altitude of p_i.

Proof. It is convenient to introduce the generating function $W_i(t,u,v_1,\dots,v_m)$ of walks having p_i as a suffix. We first show that $W, W_1,\dots,W_m$ satisfy the equation

$$WtS = W - 1 + \sum_{j=1}^{m}(v_j^{-1}-1)W_j. \tag{7}$$

To this end, we take a path $w \in \mathcal{W}$ and append a single letter $s \in \mathcal{S}$ at its end. If this produces no new occurrence of a pattern from $\mathcal{P}$, then $w.s$ is counted by $W - 1 - \sum_{j=1}^m W_j$. Otherwise, there is a new non-marked occurrence of a (uniquely determined) pattern $p_j \in \mathcal{P}$ at the end of $w.s$, and thus $w.s$ is counted by $v_j^{-1}W_j$. Now, as s can take all values in $\mathcal{S}$, this covers all the p_j's, and leads to the contribution $\sum_{j=1}^m v_j^{-1}W_j$.

[1] The notation Z, often used in statistical mechanics, is reminiscent of the word *Zustand*, which means *state* in German.

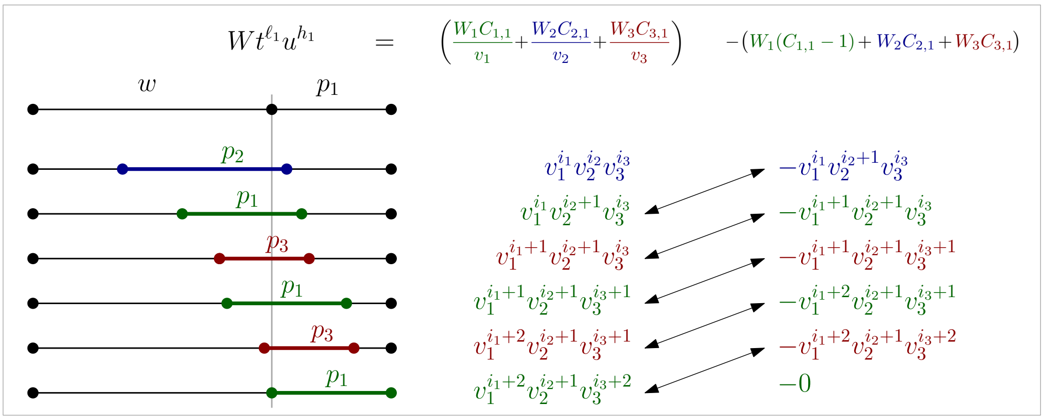
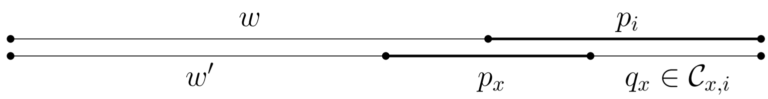


Figure 1 Illustration to the proof of Theorem 4. This schematic example (involving three patterns, p_1, p_2, and p_3) illustrates how the mutual correlation polynomials $C_{i,j}$ lead to the fact that the contributions of the decompositions $w'.p_x.q$ cancel out telescopically in the right-hand side of (8), and the full sum thus equals $v_1^{i_1}v_2^{i_2}v_3^{i_3}$, which is indeed the same as the contribution of the left-hand side of (8).

Next we show that, for each $i = 1, \ldots, m$, we have

$$Wt^{\ell_i}u^{h_i} = W_i + \sum_{j=1}^{m}(v_j^{-1} - 1)W_jC_{j,i}. \tag{8}$$

To prove this, we take a path $w \in \mathcal{W}$ and append the pattern p_i at its end, but do not mark any new occurrences of patterns from $\mathcal{P}$. For $j = 1, \ldots, m$, let γ_j be the number of occurrences of the pattern p_j in w. Then $w.p_i$ in the left-hand side of (8) contributes $v_1^{\gamma_1}v_2^{\gamma_2}\ldots v_m^{\gamma_m}$ to the generating function. Apart from p_i at the end, there are possibly some new occurrences of some patterns in $w.p_i$. Consider such a new occurrence of a pattern, say of p_x, in $w.p_i$. Then we have a decomposition $w.p_i = w'.p_x.q_x$, where $q_x \in \mathcal{C}_{x,i}$.



For $j = 1, \ldots, m$, let δ_j be the number of occurrences of the pattern p_j in $w'.p_x$. Then this decomposition contributes $v_1^{\delta_1}v_2^{\delta_2}\ldots v_m^{\delta_m}/v_x$ to $v_x^{-1}W_xC_{x,i}$, and $v_1^{\delta_1}v_2^{\delta_2}\ldots v_m^{\delta_m}$ to $W_xC_{x,i}$. Now consider the *next* new occurrence of a pattern, say of p_y, in $w.p_i$. The decomposition $w.p_i = w''.p_y.q_y$, where $q_y \in \mathcal{C}_{y,i}$, contributes $v_1^{\delta_1}v_2^{\delta_2}\ldots v_m^{\delta_m}$ to $v_y^{-1}W_yC_{y,i}$ and $v_1^{\delta_1}v_2^{\delta_2}\ldots v_m^{\delta_m}v_y$ to $W_yC_{y,i}$. Therefore, in the right-hand side of (8), the contributions of all the decompositions of $w.p_i$ will cancel out telescopically (we add W_i to cancel the contribution of $w.p_i$ itself), except the very first term whose contribution is $v_1^{\gamma_1}v_2^{\gamma_2}\ldots v_m^{\gamma_m}$. See Fig. 1 for illustration.

Finally, we regard (8) as an $m \times m$ linear system $C\mathbf{x} = \mathbf{d}$ with $\mathbf{x} = (W_1, \ldots, W_m)$ and $\mathbf{d} = (Wt^{\ell_1}u^{h_1}, \ldots, Wt^{\ell_m}u^{h_m})$. Let, further, $\mathbf{e} = (v_1^{-1}-1, \ldots, v_m^{-1}-1)$. We use the elementary fact that if $C\mathbf{x} = \mathbf{d}$ and $C^\top\mathbf{y} = \mathbf{e}$, then $\mathbf{x}\cdot\mathbf{e} = \mathbf{d}\cdot\mathbf{y}$, to find $\sum_{j=1}^{m}(v_j^{-1} - 1)W_j$. The solution $\mathbf{y}$ of $C^\top\mathbf{y} = \mathbf{e}$ can be written by Cramer's rule: we have $\mathbf{y} = \big(\det(C_1), \ldots, \det(C_m)\big)/\det(C)$, where C_i is the matrix obtained from C by replacing its ith *row* with $\mathbf{e}$. Therefore we have

$$\sum_{j=1}^{m}(v_j^{-1} - 1)W_j = \mathbf{x}\cdot\mathbf{e} = \mathbf{d}\cdot\mathbf{y} = \big(W/\det(C)\big)\sum_{i=1}^{m}t^{\ell_i}u^{h_i}\det(C_i). \tag{9}$$

We substitute this to (7), and solve for W. Finally, we multiply the numerator and the denominator by $v_1\ldots v_m$, and after some more rearrangement this yields the claimed formula (5). ◀

▶ Remark 5. Theorem 4 is a far-reaching generalization of several earlier results. For $v_1 = \ldots = v_m = 1$, we have $C = I$ and hence $\Delta = 1$, and $\Delta_i = 0$ for $1 = 1, \ldots, m$; thus $W(t, u, 1, \ldots, 1) = 1/(1-tS)$ as expected. For $v_1 = \ldots = v_m = 0$, we get the formula for walks that *avoid* $p_1, \ldots, p_m$, which was first obtained in [2]. For $m = 1$, we obtain [1, Thm. 7.1].

▶ Remark 6. Obtaining the generating function W by means of the finite automaton would generically require the inversion of an $L \times L$ matrix with symbolic coefficients, which is costly in time and in memory ($L := \sum_{i=1}^{m} \ell_i$ is the sum of the lengths of the marked patterns). It is nice that our formula based on the mutual correlation sets is algorithmically more efficient, and directly gives the generating function, avoiding those larger costs. However, comparing the two formulas for W leads to the following result (which will be used in the next section for our derivation of the closed-form formula for meanders).

▶ **Proposition 7.** *In the notation introduced above, we have*

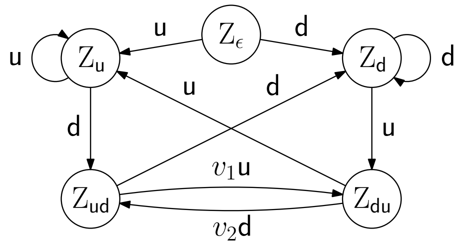
$$\Delta(t, u, v_1, \ldots, v_m) = (1, 0, \ldots, 0)\ \mathrm{adj}(I - tA)\ (1, \ldots, 1)^\top, \tag{10}$$

$$K(t, u, v_1, \ldots, v_m) := (1 - tS(u))\Delta + \sum_{i=1}^{m} \Delta_i t^{\ell_i} u^{h_i} = \det(I - tA). \tag{11}$$

Proof. Compare the formulas (4) and (5) for W, and notice that in both of them the denominator is polynomial in t with constant term 1. ◀

▶ **Definition 8.** *The expression K from* (11) *will be called the* ***kernel*** *of the walk.*

▶ **Example 9.** Consider Dyck walks (we denote the steps by $\mathsf{d} := -1$, $\mathsf{u} := 1$) with marked patterns $p_1 = \mathsf{udu}$ and $p_2 = \mathsf{dud}$. The following drawing shows the automaton for this model and its transition matrix A (the ordering of states is $Z_\epsilon, Z_\mathsf{u}, Z_\mathsf{ud}, Z_\mathsf{d}, Z_\mathsf{du}$).



$$A = \begin{pmatrix} 0 & u & 0 & u^{-1} & 0 \\ 0 & u & u^{-1} & 0 & 0 \\ 0 & 0 & 0 & u^{-1} & v_1 u \\ 0 & 0 & 0 & u^{-1} & u \\ 0 & u & v_2 u^{-1} & 0 & 0 \end{pmatrix}$$

We find $C_{1,1} = C_{2,2} = 1 + t^2, C_{1,2} = tu^{-1}, C_{2,1} = tu$; so, the mutual correlation matrix is

$$\begin{pmatrix} v_1 + (1 - v_1)(1 + t^2) & (1 - v_2)tu \\ (1 - v_1)tu^{-1} & v_2 + (1 - v_2)(1 + t^2) \end{pmatrix}.$$

By Theorem 4, we obtain the generating function for Dyck walks with marked p_1, p_2:

$$W(t, u, v_1, v_2) = \frac{1 + t^2(1 - v_1v_2) + t^4(1 - v_1)(1 - v_2)}{1 - t(u^{-1} + u) + t^2(1 - v_1v_2) - t^3\big(u^{-1}v_2(1 - v_1) + uv_1(1 - v_2)\big) - t^4(1 - v_1)(1 - v_2)}.$$

Setting v_1, v_2 to be 1 or 0, we allow or forbid the corresponding patterns. In this easy case, this recovers several known sequences, for example $W(t, u, 1, 1) = \dfrac{1}{1 - t(u^{-1} + u)}$ (as expected, since these are unrestricted walks); $W(t, 1, 0, 1) = W(t, 1, 1, 0) = \dfrac{1 + t^2}{1 - 2t + t^2 - t^3}$ (`A005251`)[2] ; $W(t, 1, 0, 0) = \dfrac{1 + t^2 + t^4}{1 - 2t + t^2 - t^4}$ (`A128588`, double Fibonacci numbers).

[2] This refers to the On-Line Encyclopedia of Integer Sequences (OEIS), available at https://oeis.org/.

▶ **Theorem 10.** *Let $\mathcal{S}$ be a set of steps, and $\mathcal{P} = \{p_1, \dots, p_m\}$ a set of (mutually not included) patterns. The* ***multivariate generating function of bridges*** *is given by*[3]

$$B(t, v_1, \dots, v_m) = \sum_{i=1}^{e} \frac{u_i'(t)}{u_i(t)} \frac{\Delta(t, u_i(t))}{K_t(t, u_i(t))}, \tag{12}$$

where $u_1(t), \dots, u_e(t)$ are the small roots of $K(t,u)$ (as defined in (11)).

Proof. To prove this formula, we extract $[u^0]$ from W, and obtain

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

via Cauchy's integral formula and the residue theorem, where the poles inside $|u| = \varepsilon$ happen to be exactly the small roots u_i. Finally, the chain rule for total derivative yields Eq. (12). ◀

▶ **Example 11.** We return to the example considered above – Dyck walks with marked $p_1 = \mathsf{udu}$ and $p_2 = \mathsf{dud}$. By Theorem 10 we obtain the generating function for bridges:

$$B(t, v_1, v_2) = \sqrt{\frac{1 + \quad (1 - v_1 v_2)t^2 + (1 - v_1)(1 - v_2)t^4}{1 + (-3 - v_1 v_2)t^2 + (1 - v_1)(1 - v_2)t^4}},$$

which, in its turn, yields sequences that appeared in earlier work in different contexts such as patterns in binary strings, but also the Potts model from statistical mechanics: $B(t,1,1) = \frac{1}{\sqrt{1-4t^2}}$ (central binomial coefficients, as expected), $B(t,0,1) = B(t,1,0) = \sqrt{\frac{1+t^2}{1-3t^2}}$ (A025565, [3, 16, 30]), $B(t,0,0) = \sqrt{\frac{1+t^2+t^4}{1-3t^2+t^4}}$ (A078678 [19, 34, 36]), etc.

3 Generating functions for meanders and excursions with marked patterns

While generating functions for walks can be found as a solution of a system of linear equations (which, in particular, implies that they are rational), the generating functions for meanders/excursions are typically algebraic (non-rational) and can be found by a suitable variation of the *kernel method*. One of them, the *vectorial kernel method*, was recently developed in [1] for dealing with enumerative problems encoded by a counter automaton. One of the cases in which this method leads to explicit formulas was that of meanders/excursions that avoid a single pattern p under the assumption that p itself is a meander. In this case, one has

$$M(t,u) = \frac{R(t,u)}{u^c K(t,u)} \prod_{i=1}^{c} (u - u_i(t)) \qquad \text{and} \qquad E(t) = -\frac{1}{t} \prod_{i=1}^{c} (-u_i(t)),$$

where c is the absolute value of the smallest number in $\mathcal{S}$, $R(t,u)$ is the autocorrelation polynomial, $K(t,u)$ is the kernel, and $u_1(t), \dots, u_c(t)$ are the small roots of $K(t,u)$. Our next theorem expands this result in two directions: first, it deals with several patterns, second, these patterns are marked and not just forbidden.

[3] Here and below we frequently remove the markers in the list of arguments of a function, writing $K(t,u)$, $\Delta(t,u)$, $u_i(t)$ for $K(t, u, v_1, \dots, v_m)$, $\Delta(t, u, v_1, \dots, v_m)$, $u_i(t, v_1, \dots, v_m)$, etc.

▶ **Theorem 12.** *Let $\mathcal{S}$ be a set of steps, and $\mathcal{P} = \{p_1, \ldots, p_m\}$ a set of (mutually not included) patterns, all of them being meanders themselves. Then the* ***multivariate generating function of meanders*** *is*

$$M(t, u, v_1, \ldots, v_m) = \frac{\Delta(t, u)}{u^c K(t, u)} \prod_{i=1}^{c} (u - u_i(t)), \tag{13}$$

where $K(t, u)$ is the kernel as in (11)*, $\Delta(t, u)$ is the determinant of the mutual correlation matrix* (6) *as in* (10)*, and $u_1(t), \ldots, u_c(t)$ are the small roots of $K(t, u)$.*
The ***multivariate generating function of excursions*** *is given by*

$$E(t, v_1, \ldots, v_m) = M(t, 0, v_1, \ldots, v_m) = -\frac{1}{t} \prod_{i=1}^{c} (-u_i(t)). \tag{14}$$

Proof. To prove (13), we apply the vectorial kernel method. According to its general scheme, we encode the meanders by the automaton, as explained before Theorem 4. We denote by M_i the generating function for meanders that terminate in state Z_i, and let $\mathbf{M} = (M_1, M_2, \ldots)$. Then we have the functional equation

$$\mathbf{M} = (1, 0, \ldots, 0) + t\mathbf{M}A - \{u^{<0}\}(t\mathbf{M}A), \tag{15}$$

where $\{u^{<0}\}(t\mathbf{M}A)$ consists of all terms of $t\mathbf{M}A$ that contain negative powers of u (in other words, $\{u^{<0}\}(t\mathbf{M}A)$ counts the paths $w.s$ such that w is a meander and $s \in \mathcal{S}$, and $w.s$ crosses the x-axis at its last step). Next we rewrite (15) as

$$\mathbf{M}(I - tA) = (1, 0, \ldots, 0) - \{u^{<0}\}(t\mathbf{M}A). \tag{16}$$

At this point we claim that in $\{u^{<0}\}(t\mathbf{M}A)$ only the first component is non-zero. This follows from the assumption that all our patterns are meanders. Therefore, if a walk $w.s$ as above has a *non-empty* overlap with $p \in \mathcal{P}$, it is impossible that its last step crosses the x-axis. This means that $w.s$ crosses the x-axis at its last step, then it is necessarily in state Z_ϵ. Therefore, negative powers of u can occur only in the first component of $t\mathbf{M}A$. Notice further that all the terms of $\{u^{<0}\}(t\mathbf{M}A)$ contain u to some powers between $-c$ and -1. Therefore we can multiply (16) by u^c and obtain

$$\mathbf{M}\, u^c\, (I - tA) = (F(t, u), 0, \ldots, 0), \tag{17}$$

where $F(t, u)$ is a monic polynomial in u of degree c.

Next we multiply (17) by $\operatorname{adj}(I - tA)(1, \ldots, 1)^\top$, and obtain, due to (10) and (11),

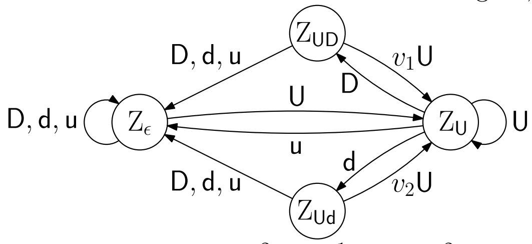
$$M(t, u)\, u^c\, K(t, u) = F(t, u)\Delta(t, u). \tag{18}$$

Here, it is legitimate to substitute, for u, any small root $u_i(t)$ of $K(t, u)$. Then the left-hand side of (18) vanishes. It is impossible that $\Delta(t, u_i(t)) = 0$ because $\Delta(t, u)$, as polynomial in t, has constant term 1 (this follows from the fact that $C_{i,i}(t, u)$ has constant term 1, and $C_{i,j}(t, u)$, $i \neq j$, has constant term 0). Therefore we have $F(t, u_i(t)) = 0$, that is, $u_i(t)$'s also are roots of $F(t, u)$.

Finally, $K(t, u)$ has precisely c small roots, $u_1(t), \ldots, u_c(t)$ (this can be proven in the same way as [1, Prop. 4.4]). Thus, $u_1(t), \ldots, u_c(t)$, are roots of $F(t, u)$, which is a monic polynomial of degree u. Therefore we know its decomposition, $F(t, u) = \prod_{i=1}^{c} (u - u_i(t))$. Now the formula (13) follows from (18).

To get (14), we substitute $u = 0$ and notice that the only term in the denominator that does not vanish is $-t\Delta$. ◀

▶ **Example 13.** Basketball walks are lattice paths with $\mathcal{S}=\{-2,-1,1,2\}$. We also denote their steps by $\mathsf{D}=-2$, $\mathsf{d}=-1$, $\mathsf{u}=1$, $\mathsf{U}=2$. In this example we find the generating functions for meanders and excursions with marked $p_1=\mathsf{UDU}$ and $p_2=\mathsf{UdU}$. The automaton and its transition matrix are shown in the next figure, the ordering of the states is $\mathrm{Z}_\epsilon, \mathrm{Z}_{\mathsf{U}}, \mathrm{Z}_{\mathsf{UD}}, \mathrm{Z}_{\mathsf{Ud}}$.



$$A=\begin{pmatrix} u^{-2}+u^{-1}+u & u^2 & 0 & 0\\ u & u^2 & u^{-2} & u^{-1}\\ u^{-2}+u^{-1}+u & v_1u^2 & 0 & 0\\ u^{-2}+u^{-1}+u & v_2u^2 & 0 & 0\end{pmatrix}$$

We have $S(u)=u^{-2}+u^{-1}+u+u^2$ and $c=2$. The mutual correlation polynomials are $C_{11}=1+t^2$, $C_{12}=t^2u$, $C_{21}=t^2$, $C_{22}=1+t^2u$. By Theorem 4, we obtain $\Delta=1+t^2+t^2u$ and $K=-\big((t+t^3)+(t+2t^3)u-(1+t^2-t^3)u^2+(t-t^2+t^3)u^3+(t+t^3)u^4\big)/u^2$. Thus, u^2K is a polynomial of degree 4 with two small roots given by given by Puiseux series

$$u_{1,2}(t)=\pm t^{\frac12}+\frac12 t\pm\frac18 t^{\frac32}+\frac12 t^2\pm\frac{159}{128}t^{\frac52}+\frac32 t^3\pm\frac{1761}{1024}t^{\frac72}+\frac72 t^4\pm\frac{213435+16384v_1}{32768}t^{\frac92}+\frac{19+2v_1+v_2}{2}t^5\pm\dots$$

By Theorem 12, we obtain generating functions for meanders/excursions with *marked* p_1,p_2:

$$M(t,u,v_1,v_2)=\frac{\Delta(t,u)}{u^2K(t,u)}(u-u_1(t))(u-u_2(t))=$$
$$1+(u+u^2)t+(2+u+u^2+2u^3+u^4)t^2+(2+5u+(5+v_1)u^2+(2+v_2)u^3+3u^4+3u^5+u^6)t^3+\dots,$$
$$E(t,v_1,v_2)=\frac{u_1(t)u_2(t)}{-t}=1+2t^2+2t^3+(10+v_1)t^4+(21+v_1+2v_2)t^5+(79+9v_1+4v_2+v_1^2)t^6+\dots$$

(For example, there is one excursion of size 5 that contains UDU, namely UDUdd, and two excursions that contain UdU, namely UdUdD, and UdUDd.)

To obtain the univariate generating functions for all meanders and that for excursions that *avoid* p_1,p_2, we substitute $v_1=v_2=0$, and $u=1$ resp. $u=0$:

$$M(t)=1+2t+7t^2+21t^3+71t^4+245t^5+867t^6+3091t^7+11147t^8+40491t^9+148010t^{10}+\dots,$$
$$E(t)=1+2t^2+2t^3+10t^4+21t^5+79t^6+224t^7+771t^8+2462t^9+8409t^{10}+\dots$$

▶ Remark 14. If some of the patterns are not meanders, then generically several components of $\{u^{<0}\}(t\mathbf{M}A)$ are non-zero. Therefore, in general one does not get simple equations as (17) and (18), and the formula (13) does not hold verbatim. However, it is then possible to use the approach introduced in [1, Thm. 3.2]; this gives that $M(t,u)$ has the form $\frac{G(t,u)}{u^cK(t,u)}\prod_{i=1}^c(u-u_i(t))$, where $G(t,u)$ is polynomial in u. There are other cases, not covered by Theorem 12, where it is possible to find formulas for $M(t,u)$ and $E(t,u)$. For example, if the only negative step in $\mathcal{S}$ is -1 (such paths are called *Łukasiewicz walks*), one can use the fact that a path can cross the x-axis only when a (-1)-step is appended to an excursion. Using further ideas developed (for avoidance) in [2], we can find, for example, generating functions for Dyck meanders/excursions with marked $p_1=\mathsf{udu}, p_2=\mathsf{dud}$:

$$M(t,u,v_1,v_2)=\left(1-\frac{t^2(1-v_1)(1-v_2)}{2}\left(1-\sqrt{1-4t^2/\Delta}\right)\right)\frac{\Delta}{uK}(u-u_1(t)),$$
$$E(t,v_1,v_2)=\frac{\Delta}{1+t^2v_2(1-v_1)+t^3(1-v_1)(1-v_2)u_1(t)}\frac{u_1(t)}{t}=\frac{\Delta}{2t^2}\left(1-\sqrt{1-\frac{4t^2}{\Delta}}\right),$$

where $\Delta=1+t^2(1-v_1v_2)+t^4(1-v_1)(1-v_2)$ as found in Example 9 (note that another form of $E(t,v_1,v_2)$ is mentioned in A145895).

4 Multivariate Gaussian limit laws for pattern occurrences

4.1 Gaussian and multivariate Gaussian distribution

The Gaussian distribution is ubiquitous is mathematics, physics, biology, astronomy, finance, computer science, and even in human sciences, and, in fact, in any domain in which one could collect numerical data and do some statistics with them.

There are two frequent simple explanations of this universality.

- The first explanation is probabilistic: the central limit theorem of Laplace asserts that if one considers a sequence of independent and identically distributed random variables $(X_n)_{n\in\mathbb{N}}$ (with expected value μ and finite variance σ^2), then the sum $\sum_{k=1}^{n} X_k$ is converging towards the Gaussian distribution $\mathcal{N}(\mu, \sigma)$.
- The second explanation is analytic: if the corresponding probability generating function of $\mathrm{Prob}(X_n = k)$ behaves like a "quasi-power" (see [29]), then X_n has a Gaussian limit distribution.

Both approaches have their own interest, as both admit some flexibility in their condition of application. As masterfully presented by Flajolet and Sedgewick in [22], the second approach is typically split into two steps: first a combinatorial step consists in getting a closed-form expression (or a functional equation) for the generating function, and then a local analysis of this function near its dominant singularity is performed in order to get some universal behaviour (limit law, etc).

We apply a generalization of this analytico-combinatorial approach to the case of joint laws $\Pr(X_1 = k_1, \ldots, X_m = k_m)$, including in cases where the random variables X_i are dependent. The dependence (the correlations) will be handled at the level of the generating function, on which some multivariate complex analysis is then performed in order to get the limit law. As a first step towards more examples in the realm of "multivariate analytic combinatorics" (as initiated in [12, 20, 28, 37]), we present here some results related to the multivariate Gaussian distribution.

For the tuple of random variables $\boldsymbol{X} = (X_1, \ldots, X_m)$ of average $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ the associated covariance matrix $\boldsymbol{\Sigma}$ is defined by

$$\Sigma_{ij} := \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \qquad (\text{for } i, j = 1, \ldots, m). \tag{19}$$

This matrix $\boldsymbol{\Sigma}$ is also sometimes called the variance-covariance matrix, as the diagonal terms are exactly the variance of each X_i. Note that $\boldsymbol{\Sigma}$ is a positive-definite matrix, therefore $\sqrt{\det \boldsymbol{\Sigma}}$ is well defined.

The multivariate Gaussian distribution (also called multivariate normal distribution, or m-dimensional Gaussian distribution, see e.g. [17]), denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, is a generalization of the classical (one-dimensional) normal distribution; its density is

$$\frac{1}{\sqrt{(2\pi)^m \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right). \tag{20}$$

When all the bold quantities are scalars (i.e. when $m = 1$), it is coinciding with the classical expression for the density of the Gaussian distribution.

Let us now illustrate this multivariate approach on fundamental objects such as words and constrained lattice paths. We first present a nice unifying example, before switching to more general cases from the algebraic world.

4.2 A multi-multivariate generating function for all patterns at once

There is a vast amount of literature on Dyck, Motzkin, Schröder etc. lattice paths (or some related classes of RNA structures, ordered trees, permutations) in which some combination of patterns (valleys, peaks, etc.) are considered. Proofs of such results often rely on some ad hoc context-free grammar decompositions; see e.g. [10, 15, 18, 21, 31, 33]. The power of our approach is also in the fact that it enables us to obtain many such results *at once* by marking sufficiently many patterns and then setting them to be 0 or 1 in any desirable combinations.

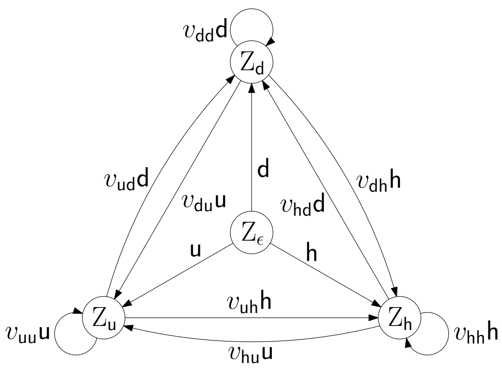
We illustrate this for the model of Motzkin walks ($\mathcal{S} = \{\mathsf{d} = -1, \mathsf{h} = 0, \mathsf{u} = 1\}$) in which we mark all possible patterns of length 2. To this aim, we introduce nine markers for all such patterns (v_{ud} for the pattern ud, etc.), and we obtain an even more explicit formula in cases not covered by the closed-form formula from Theorem 12. We give the general expression for excursions in the following theorem (the general expression for meanders is somewhat more lengthy, see also [2, Thm. 4]).

▶ **Theorem 15.** *The generating function $E(t, v_{\mathsf{uu}}, v_{\mathsf{uh}}, v_{\mathsf{ud}}, v_{\mathsf{hu}}, v_{\mathsf{hh}}, v_{\mathsf{hd}}, v_{\mathsf{du}}, v_{\mathsf{dh}}, v_{\mathsf{dd}})$ of Motzkin excursions, where each v_p counts the number of occurrences of the pattern p, is*

$$\frac{(v_{\mathsf{dd}} - 1) - t((v_{\mathsf{dd}} - 1)v_{\mathsf{hh}} - (v_{\mathsf{dh}} - 1)v_{\mathsf{hd}} - v_{\mathsf{dd}} + v_{\mathsf{dh}}) + \big(1 + t(v_{\mathsf{dh}} - v_{\mathsf{hh}})\big)\frac{u\mathbf{v}_1}{t\mathbf{v}_4}\big|_{u=u_1(t)}}{v_{\mathsf{dd}} + t(v_{\mathsf{dh}}v_{\mathsf{hd}} - v_{\mathsf{dd}}v_{\mathsf{hh}})}, \qquad (21)$$

where $u_1(t)$ is the unique small solution of the kernel $K(t, u)$, and $\mathbf{v}_1$ and $\mathbf{v}_4$ are the 1st and the 4th components of $\mathbf{v} := \operatorname{adj}(I - tA)(1, \ldots, 1)^\top$.

Proof (Sketch). This model is encoded by the following automaton and its transition matrix:

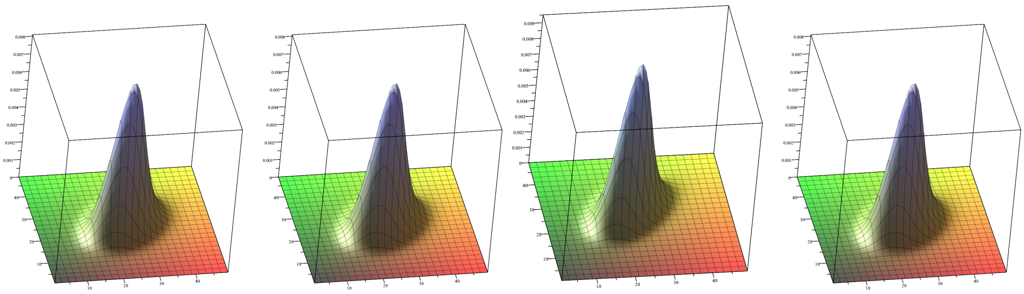


$$\begin{array}{cc} & \begin{array}{cccc} Z_\epsilon & Z_{\mathsf{u}} & Z_{\mathsf{h}} & Z_{\mathsf{d}} \end{array} \\ A = & \begin{pmatrix} 0 & u & 1 & u^{-1} \\ 0 & v_{\mathsf{uu}}u & v_{\mathsf{uh}} & v_{\mathsf{ud}}u^{-1} \\ 0 & v_{\mathsf{hu}}u & v_{\mathsf{hh}} & v_{\mathsf{hd}}u^{-1} \\ 0 & v_{\mathsf{du}}u & v_{\mathsf{dh}} & v_{\mathsf{dd}}u^{-1} \end{pmatrix} \end{array}$$

A Motzkin path can cross the x axis only by reading d (that is, entering the 4th state). Thus, only the fourth component of $t\mathbf{M}A$ has terms with negative powers of u. This leads to the equation $\mathbf{v}_1(t, u) - \mathbf{v}_4(t, u)N(t, u) = 0$ where $N(t, u)$ is the generating function for the terms with negative powers of u in the fourth component of $t\mathbf{M}A$. Note that by analyzing which patterns are read if $w.s$ crosses the x-axis, we can express N in terms of E.

Finally, as $uK(t, u)$ is a polynomial of degree 2 in u, it has one small root, $u_1(t)$ (we dropped the dependency on the other variables v_p's of the kernel (11)). Now, by the vectorial kernel method (see the proof of Theorem 12), this leads to the formula (21) for E. ◀

Setting the markers v_p to be 1 or 0 in all possible combinations leads to 512 specific models. An exhaustive analysis shows that they lead to 75 distinct sequences for excursions and 158 distinct sequences for meanders. In some cases we obtain new interpretations for existing OEIS entries, thus potentially leading to new bijections between different combinatorial structures.



walks $\mathbb{E}[X_i(n)] = \frac{n}{8} - \frac{1}{4}$ | bridges $\mathbb{E}[X_i(n)] \sim \frac{n}{8} - \frac{1}{8}$ | meanders $\mathbb{E}[X_i(n)] \sim \frac{n}{8} \pm \frac{\sqrt{2\pi n}}{16}$ | excursions $\mathbb{E}[X_i(n)] \sim \frac{n}{8} + \frac{1}{8}$

Figure 2 Distribution (X_1, X_2) of the pair of patterns $(\mathsf{udu}, \mathsf{dud})$ in a Dyck walks/bridges/meanders/excursions of length $n = 200$ (this corresponds to the model of Example 9). Already, for this small value of n, one sees that $\text{Prob}(X_1 = k_1, X_2 = k_2)$ is concentrated around the value $(\mathbb{E}[X_1], \mathbb{E}[X_2])$ with Gaussian fluctuations. (This example has by design a symmetric behaviour for X_1 and X_2 for walks, bridges, and excursions; this is not generically the case.)

Moreover, we checked that all these models satisfy the technical conditions (see [28, 37]) which ensure a *multivariate Gaussian distribution*. We now discuss more general models.

First, let us mention that the case of walks without positivity constraint or final altitude constraint is easier: indeed, their generating function W is rational, and one can then more directly apply results from [28, 36] to get the multivariate Gaussian distribution. Note that if one allows to mark a regular expression (and not just a finite set of words), then, already in the rational case, one can get "any" arbitrary (non-Gaussian) distribution (see [4] for a presentation of this huge diversity of the possible limit laws for pattern occurrences). It is more involved to analyse the algebraic generating function cases; one can however still prove that the multivariate Gaussian distribution also holds (see Figure 2 for an illustration):

▶ **Theorem 16.** *For any generic model of walks, let $X_i(n)$ be the random variable counting occurrences of the pattern p_i (for $i = 1, \ldots, m$) in a bridge/excursion/meander of length n. Then the joint law $(X_1(n), \ldots, X_m(n))$ convergences to a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as defined in Section 4.1.*

Proof (Sketch). Some technical conditions are required to avoid degenerated cases: for lattice paths, this corresponds to what is called *generic* model of walks in [1, Definition 6.1]; this definition includes conditions like having a unique dominant singularity, that the number of paths of length n is strictly increasing for large n, etc. Then, all the univariate asymptotics follow the universal asymptotics established in [1, 5, 6].

Now, the multivariate asymptotics follow the algebraic schemes investigated in [23, 26], and thus lead to the multivariate Gaussian distribution. It is also possible to use a multivariate Spitzer/Sparre Andersen formula (see [8, Theorem 8]), rephrased as

$$W^+(t, u, v_1, \ldots, v_m) := \{u^{\geqslant 0}\} W(t, u, v_1, \ldots, v_m) = [s^0] W(t, su, v_1, \ldots, v_m) 1/(1 - 1/s),$$

$$M(t, u, v_1, \ldots, v_m) \sim \exp \int_0^t \frac{W^+(z, u, v_1, \ldots, v_m) - 1}{z} dz. \tag{22}$$

In fact, Formula (22) is an equality when $v_i = 1$ (for $i = 1, \ldots, m$), while, if one keeps track of the v_i's, the counting of occurrences of p_i (in a meander of length n) could differ by a few $O(1)$ occurrences between both sides of Formula (22): indeed, the proof uses a concatenation of some final and initial parts of the path, and this can create/delete a few occurrences of p_i's.

The advantage of using the multivariate Spitzer formula is that this relates the meanders to a diagonal involving the rational generating function of walks, on which one can apply the results of [36]; the drawback is that one loses asymptotics below the $O(1)$ precision.

Let us now state how to derive the parameters of the multidimensional Gaussian limit law $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $F(t, u, v_1, \ldots, v_m)$ be the corresponding generating function (where t encodes the length, u the final altitude, and each v_i encodes occurrences of the pattern p_i). The average of the marginals behaves linearly, as expected by the Borges theorem (see [1,22]):

$$\mathbb{E}[X_i(n)] = \frac{[t^n]\partial_{v_i}(F)(t, 1, \ldots, 1)}{[t^n]F(t, 1, \ldots, 1)} = \mu_i n(1 + o(1)). \tag{23}$$

Note that, in (23), there would be no difficulty in pushing the asymptotics further than $o(1)$. One sets $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_m)$. Now, the entries of the covariance matrix $\boldsymbol{\Sigma}$ are obtained by

$$\begin{aligned}\boldsymbol{\Sigma}_{ij} &= \lim_{n\to\infty} \frac{1}{n^2}\mathbb{E}\left[(X_i(n) - \mathbb{E}[X_i(n)])\,(X_j(n) - \mathbb{E}[X_j(n)])\right] \\ &= \lim_{n\to\infty} \frac{1}{n^2} \frac{[t^n]\partial_{v_i}\partial_{v_j}(F)(t, 1, \ldots, 1)}{[t^n]F(t, 1, \ldots, 1)} - \mu_i\mu_j. \end{aligned} \tag{24}$$

One has $\boldsymbol{\Sigma}_{ij} > 0$ as a consequence of the universal positivity of the variability condition [8, Lemma 22]) and $\det \boldsymbol{\Sigma} \neq 0$ when the patterns p_i's are not all equal.

Thus, when one considers the asymptotic regime of $[z^n v_1^{\mu_1 n} \ldots v_m^{\mu_m n}]F(z, v_1, \ldots, v_m)$, where the exponents of the v_i's can be rounded to the nearest integer whenever needed, one gets an expansion which fits the framework of the multivariate version of the quasi-power theorem (see [28,29,37]), leading to the multidimensional Gaussian limit law $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. ◀

▶ Remark. Generic walks are aperiodic; a multivariate Gaussian distribution is also holding for periodic walks, but additional care is required. (Recall that a walk is periodic if the gcd of the differences between the steps of $\mathcal{S}$ is not 1, and then the paths live in a periodic sub-lattice of $\mathbb{Z}^2$, and then the generating function has conjugate dominant singularities.) These periodic cases can in fact be handled by combining the approaches of [5] and [9].

5 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 functions of constrained lattice paths (walks and bridges in Section 2, meanders and excursions in Section 3), generalizing the previous works [1,2,6] and leading to multidimensional Gaussian limit laws.

This will allow us to tackle further questions, like

- faster uniform random generation of constrained paths of length n, extending the multivariate tuning of the Boltzmann method done in [14] to cases where the grammar is not strongly connected (such cases are generic for lattice paths with forbidden patterns),
- links with trace monoids and partial commutations in words [13],
- to extend the analysis of fundamental non-Gaussian parameters under some pattern constraints (like the area below the path [7] for walks with forbidden patterns, thus interfering with the natural drift of the walk), possibly combined with further constraints (like to be below a line of rational slope, extending [9]).

This work is also a first step towards more general schemes of multidimensional limit laws in analytic combinatorics, for the important class of algebraic functions related to lattice path statistics.

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*, 82(3):386–428, 2020. doi:10.1007/s00453-019-00623-3.
2. Andrei Asinowski, Cyril Banderier, and Valerie Roitner. Generating functions for lattice paths with several forbidden patterns. In *Proceedings of Formal Power Series and Algebraic Combibatorics (FPSAC'20)*, Ramat-Gan, Israel, 2020. To appear.
3. Davide Baccherini, Donatella Merlini, and Renzo Sprugnoli. Binary words excluding a pattern and proper Riordan arrays. *Discrete Mathematics*, 307(9):1021–1037, 2007. doi:10.1016/j.disc.2006.07.023.
4. Cyril Banderier, Olivier Bodini, Yann Ponty, and Hanane Tafat Bouzid. Biodiversity of pattern distributions in combinatorial ecosystems. In *The SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO'12)*, pages 107–115, Kyoto, Japan, January 2012. doi:10.1137/1.9781611973020.13.
5. Cyril Banderier and Michael Drmota. Formulae and asymptotics for coefficients of algebraic functions. *Combinatorics, Probability and Computing*, 24(1):1–53, 2015. doi:10.1017/S0963548314000728.
6. Cyril Banderier and Philippe Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science*, 281(1-2):37–80, 2002. doi:10.1016/S0304-3975(02)00007-5.
7. Cyril Banderier and Bernhard Gittenberger. Analytic combinatorics of lattice paths: enumeration and asymptotics for the area. *Discrete Math. Theor. Comput. Sci. Proc.*, AG:345–355, 2006. Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities, 2006, Nancy, France. URL: https://dmtcs.episciences.org/3481.
8. Cyril Banderier, Marie-Louise Lackner, and Michael Wallner. Latticepathology and symmetric functions. In *31th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020)*, volume 159 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:16, 2020. doi:10.4230/LIPIcs.AofA.2020.2.
9. Cyril Banderier and Michael Wallner. The kernel method for lattice paths below a rational slope. In *Lattice paths combinatorics and applications*, Developments in Mathematics Series, pages 119–154. Springer, 2019. doi:10.1007/978-3-030-11102-1_7.
10. Jean-Luc Baril, Sergey Kirgizov, and Armen Petrossian. Enumeration of Łukasiewicz paths modulo some patterns. *Discrete Math.*, 342(4):997–1005, 2019. doi:10.1016/j.disc.2018.12.005.
11. Frédérique Bassino, Julien Clément, and Pierre Nicodème. Counting occurrences for a finite set of words: Combinatorial methods. *ACM Trans. Algorithms*, 8(3):31:1–31:28, 2012. doi:10.1145/2229163.2229175.
12. Edward A. Bender and L. Bruce Richmond. Central and local limit theorems applied to asymptotic enumeration. II. Multivariate generating functions. *J. Combin. Theory Ser. A*, 34(3):255–265, 1983. doi:10.1016/0097-3165(83)90062-6.
13. 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. doi:10.1016/S0304-3975(02)00892-7.
14. Olivier Bodini and Yann Ponty. Multi-dimensional Boltzmann sampling of languages. *Disc. Math. Theor. Comp. Sci.*, AM, January 2010. URL: https://dmtcs.episciences.org/2793.
15. Charlotte Brennan and Simon Mavhungu. Visits to level r by Dyck paths. *Fund. Inform.*, 117(1-4):127–145, 2012. URL: https://www.researchgate.net/publication/264970113_Visits_to_Level_r_by_Dyck_Paths.
16. Shu-Chiuan Chang and Robert Shrock. Structural properties of Potts model partition functions and chromatic polynomials for lattice strips. *Physica A: Statistical Mechanics and its Applications*, 296(1–2):132–182, 2001. doi:10.1016/S0378-4371(01)00150-9.

17 Christopher Chatfield and Alexander J. Collins. *Introduction to Multivariate Analysis*. Chapman & Hall, London-New York, 1980. doi:10.1007/978-1-4899-3184-9.

18 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. doi:10.1016/S0012-365X(02)00341-2.

19 Tomislav Doslic. Seven (lattice) paths to log-convexity. *Acta Applicandae Mathematicae*, 110:1373–1392, 2010. doi:10.1007/s10440-009-9515-4.

20 Michael Drmota, Bernhard Gittenberger, and Johannes F. Morgenbesser. Infinite systems of functional equations and Gaussian limiting distributions. *Discrete Math. Theor. Comput. Sci. Proc., AQ*, pages 453–478, 2015. 23rd International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'12), 2012, Montreal, Canada. URL: https://dmtcs.episciences.org/3012.

21 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. doi:10.1111/1467-9590.t01-1-00042.

22 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: http://algo.inria.fr/flajolet/Publications/book.pdf.

23 Zhicheng Gao and L. Bruce Richmond. Central and local limit theorems applied to asymptotic enumeration. IV: Multivariate generating functions. *J. Comput. Appl. Math.*, 41(1-2):177–186, 1992. doi:10.1016/0377-0427(92)90247-U.

24 Ian P. Goulden and David M. Jackson. *Combinatorial enumeration*. Wiley, 1983.

25 Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics*. Addison-Wesley Publishing Company, 2nd edition, 1994. (1st edition: 1989).

26 Torin Greenwood. Asymptotics of bivariate analytic functions with algebraic singularities. *J. Comb. Theory, Ser. A*, 153:1–30, 2018. doi:10.1016/j.jcta.2017.06.014.

27 Leonidas J. Guibas and Andrew M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, 30(2):183–208, 1981. doi:10.1016/0097-3165(81)90005-4.

28 Clemens Heuberger and Sara Kropf. Higher dimensional quasi-power theorem and Berry-Esseen inequality. *Monatsh. Math.*, 187(2):293–314, 2018. doi:10.1007/s00605-018-1215-6.

29 Hsien-Kuei Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European J. Combin.*, 19(3):329–343, 1998. doi:10.1006/eujc.1997.0179.

30 Jesper Jacobsen and Jesús Salas. Transfer matrices and partition-function zeros for antiferromagnetic potts models. IV. Chromatic polynomial with cyclic boundary conditions. *Journal of Statistical Physics*, 122:705–760, 2006. doi:10.1007/s10955-005-8077-8.

31 Emma Y. Jin, Jing Qin, and Christian M. Reidys. Combinatorics of RNA structures with pseudoknots. *Bull. Math. Biol.*, 70(1):45–67, 2008. doi:10.1007/s11538-007-9240-y.

32 Yong Kong. Extension of Goulden–Jackson cluster method on pattern occurrences in random sequences and comparison with Régnier–Szpankowski method. *J. Difference Equ. Appl.*, 11(15):1265–1271, 2005. doi:10.1080/10236190500376326.

33 Toufik Mansour and Mark Shattuck. Counting humps and peaks in generalized Motzkin paths. *Discrete Appl. Math.*, 161(13-14):2213–2216, 2013. doi:10.1016/j.dam.2013.03.007.

34 Emanuele Munarini and Norma Zagaglia Salvi. Binary strings without zigzags. *Sém. Lothar. Combin.*, 49:Article B49h, 15 pp., 2002. URL: https://www.mat.univie.ac.at/~slc/wpapers/s49zagaglia.pdf.

35 John Noonan and Doron Zeilberger. The Goulden–Jackson cluster method: Extensions, applications and implementations. *J. Difference Equ. Appl.*, 5(4-5):355–377, 1999. doi:10.1080/10236199908808197.

36 Robin Pemantle and Mark C. Wilson. Twenty combinatorial examples of asymptotics derived from multivariate generating functions. *SIAM Rev.*, 50(2):199–272, 2008. URL: https://arxiv.org/pdf/math/0512548.

37 Robin Pemantle and Mark C. Wilson. *Analytic Combinatorics in Several Variables*. Cambridge University Press, 2013. URL: `https://www.math.upenn.edu/~pemantle/ACSV.pdf`.

38 Marcel-Paul Schützenberger. On the synchronizing properties of certain prefix codes. *Information and Control*, 7:23–36, 1964. URL: `https://www.sciencedirect.com/science/article/pii/S0019995864902323/pdf`.

39 Yan Zhuang. A generalized Goulden–Jackson cluster method and lattice path enumeration. *Discrete Math.*, 341(2):358–379, 2018. `doi:10.1016/j.disc.2017.09.004`.