Analytic Combinatorics of Lattice Paths with Forbidden Patterns: Asymptotic Aspects and Borges's Theorem

Andrei Asinowski
Technische Universität Wien, Institut für Diskrete Mathematik und Geometrie,
`https://dmg.tuwien.ac.at/asinowski/`
https://orcid.org/0000-0002-0689-0775

Axel Bacher
Université Paris 13, LIPN, UMR CNRS 7030, `http://lipn.univ-paris13.fr/~bacher`
https://orcid.org/0000-0002-9789-7074

Cyril Banderier
Université Paris 13, LIPN, UMR CNRS 7030, `http://lipn.univ-paris13.fr/~banderier`
https://orcid.org/0000-0003-0755-3022

Bernhard Gittenberger
Technische Universität Wien, Institut für Diskrete Mathematik und Geometrie,
`https://dmg.tuwien.ac.at/gittenberger/`
https://orcid.org/0000-0002-2639-8227

Abstract

In a companion article dedicated to the enumeration aspects, we showed how to obtain closed form formulas for the generating functions of walks, bridges, meanders, and excursions avoiding any fixed word (a *pattern* p). The autocorrelation polynomial of this forbidden pattern p (as introduced by Guibas and Odlyzko in 1981, in the context of regular expressions) plays a crucial role. In this article, we get the asymptotics of these walks. We also introduce a trivariate generating function (length, final altitude, number of occurrences of p), for which we derive a closed form. We prove that the number of occurrences of p is normally distributed: This is what Flajolet and Sedgewick call an instance of Borges's theorem.

We thus extend and refine the study by Banderier and Flajolet in 2002 on lattice paths, and we unify several dozens of articles which investigated patterns like peaks, valleys, humps, etc., in Dyck and Motzkin paths. Our approach relies on methods of analytic combinatorics, and on a matricial generalization of the kernel method. The situation is much more involved than in the Banderier–Flajolet work: forbidden patterns lead to a wider zoology of asymptotic behaviours, and we classify them according to the geometry of a Newton polygon associated with these constrained walks, and we analyse what are the universal phenomena common to all these models of lattice paths avoiding a pattern.

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 paths, pattern avoidance, finite automata, context-free languages, autocorrelation, generating function, kernel method, asymptotic analysis, Gaussian limit law

Digital Object Identifier 10.4230/LIPIcs.AofA.2018.10

Funding This work was partially funded by the Austrian Science Fund (FWF) project SFB F50 "Algorithmic and Enumerative Combinatorics".

Acknowledgements We thank our referees for their careful reading.

29th International Conference on Probabilistic, Combinatorial andAsymptotic Methods for the Analysis of Algorithms.
Editors: James Allen Fill and Mark Daniel Ward; Article No. 10; pp. 10:1–10:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

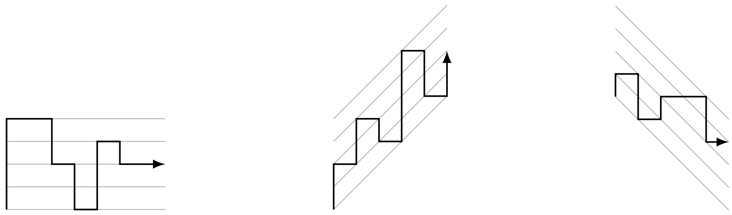


Figure 1 Some models of self-avoiding walks are encoded by partially directed lattice paths avoiding a pattern (see [2]). In this article, we analyse the asymptotics of more general walks (any jumps, any forbidden pattern).

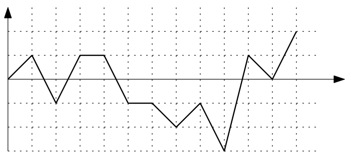
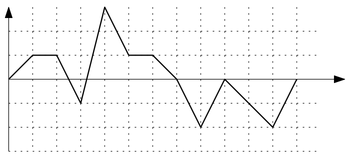
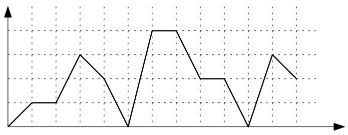
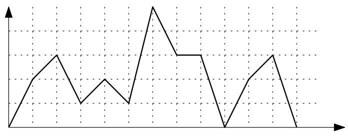
1 Introduction

Combinatorial structures having a rational or an algebraic generating function play a key role in many fields: computer science (analysis of algorithms involving trees, lists, words), computational geometry (integer points in polytopes, maps, graph decomposition), bioinformatics (RNA structure, pattern matching), number theory (integer compositions, automatic sequences and modular properties, integer solutions of varieties), probability theory (Markov chains, directed random walks), see e.g. [4, 11, 22, 32]. They are often the trace of a structure which has a recursive specification in terms of a system of tree-like structures, or of some functional equation solvable by variants of the kernel method [12].

Since the seminal article by Chomsky and Schützenberger on the link between context-free grammars and algebraic functions [15], which also holds for pushdown automata [30], many articles encoded and enumerated combinatorial structures via a formal language approach. See e.g. [20, 25, 28] for such an approach on the so-called *generalized Dyck languages*. The words generated by these languages are in bijection with directed lattice paths, and in this article, we try to understand how some of these fundamental objects can be enumerated when they have the additional constraint to avoid a given pattern. For sure, such a class of objects can be described as the intersection of a context-free language and a rational language; therefore, classical closure properties imply that they are directly generated by another (but huge and clumsy) context-free language. Unfortunately, despite the fact that the algebraic system associated with the corresponding context-free grammar is *in theory* solvable by a resultant computation or by Gröbner bases, this leads *in practice* to equations which are so big that no current computer could handle them in memory, even for generalized Dyck languages with only 20 different letters.

In this article, we generalize the asymptotics obtained by Banderier and Flajolet [5] to lattice paths avoiding a given pattern. As we shall see, the situation is much more involved, and we build on the explicit formulas that we obtained in our companion article [1]. There, we introduced a generic way to tackle the question of enumerating words avoiding a given pattern (for languages generated by pushdown automata) which bypass these intractable equations. For directed lattice paths, our method allows to handle an arbitrary number of letters (*i.e.*, allowed steps), up to alphabets of thousands of letters, computationally in a few minutes. It relies on an analytic combinatorics approach, and also on the kernel method, which we used in our investigation of enumerative and asymptotic properties of lattice paths [6–8]. This allows to unify the considerations of many articles which investigated natural patterns like peaks, valleys, humps, *etc.*, in Dyck and Motzkin words, corresponding patterns in trees, compositions..., see e.g. [9, 10, 13, 16, 17, 19, 21, 26, 27, 29].

Table 1 Summary of our results from [1], which extend the Banderier–Flajolet results from [5] to lattice paths avoiding a pattern. For the four types of paths and for any set of jumps encoded by $P(u)$, we give the corresponding generating function of such lattice paths avoiding a pattern p (of length ℓ and final altitude b). The formulas involve the autocorrelation polynomial $R(t,u)$ of p, and the small roots u_i of the kernel $K(t,u) := (1-tP(u))R(t,u)+t^\ell u^b$.

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

2 Definitions and notations

In their paper, Banderier and Flajolet consider the following setting. 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 $w=[v_1,v_2,\ldots,v_n]$ 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 $(1,v_1),(1,v_2),\ldots,(1,v_n)$. The letters that form the path $w=[v_1,v_2,\ldots,v_n]$ are referred to as its *steps*. The *length* of w, to be denoted by $|w|$, 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 $v_1+v_2+\ldots+v_n$.

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.

In the generating functions, the variable t corresponds to the length of a path, and the variable u to its final altitude. $P(u)$ is the *characteristic polynomial* of the set of steps $\mathcal{S}$, defined by $P(u)=\sum_{s\in\mathcal{S}}u^s$. The smallest (negative) number in $\mathcal{S}$ is denoted by $-c$, and the largest (positive) number in $\mathcal{S}$ is denoted by d: that is[1], if one orders the terms of $P(u)$ by the powers of u, one has $P(u)=u^{-c}+u^{s_2}+\ldots+u^{s_{|\mathcal{S}|-1}}+u^d$ with $\mathcal{S}=\{-c,s_2,\ldots,s_{|\mathcal{S}|-1},d\}$. The drift of the walk is given by the quantity $P'(1)$.

[1] Some weights (or probabilities, or multiplicities) could be associated with each jump, but we omit them in this article for clarity. All the proofs would be similar.

3 Lattice paths with forbidden patterns and the autocorrelation polynomial

We consider lattice paths with step set $\mathcal{S}$ that avoid a certain *pattern*, that is, an *a priori* fixed path $p = [a_1, a_2, \ldots, a_\ell]$. To be precise, we define an *occurrence* of p in a lattice path w as a substring of w which coincides with p. If there is no occurrence of p in w, we say that w *avoids* p. For example, the path $[1, 2, 3, 1, 2]$ has two occurrences of $[1, 2]$, but it avoids $[2, 1]$.

Before we state our results, we introduce some notations.

A *presuffix* of p is a non-empty string that occurs in p both as a prefix and as a suffix. In particular, the whole word p is a (trivial) presuffix of itself. If p has one or several non-trivial presuffixes, we say that p exhibits an *autocorrelation* phenomenon. For example, for the pattern $p = [1, 1, 2, 1, 2]$ we have no autocorrelation. In contrast, the pattern $p = [1, 1, 2, 3, 1, 1, 2, 3, 1, 1]$ has three non-trivial presuffixes: $[1]$, $[1, 1]$, and $[1, 1, 2, 3, 1, 1]$, and thus in this case we have autocorrelation.

While analysing the Boyer–Moore string searching algorithm and properties of periodic words, Guibas and Odlyzko [23] introduced in 1981 what turns out to be one of the key characters of our article, the autocorrelation polynomial[2] of the pattern p: For any given word p, let $\mathcal{Q}$ be the set of its presuffixes; the *autocorrelation polynomial* of p is

$$R(t, u) = \sum_{q \in \mathcal{Q}} t^{|\bar{q}|} u^{h(\bar{q})}, \tag{1}$$

where $\bar{q}$ denotes the complement of q in p (i.e. $q\bar{q} = p$) and $h(\bar{q})$ the final altitude of a walk made of the steps of $\bar{q}$.

For example, consider the pattern $p = [1, 1, 2, 3, 1, 1, 2, 3, 1, 1]$. Its four presuffixes produce four terms of $R(t, u)$ as follows:

q	$\lvert\bar{q}\rvert$	$h(\bar{q})$
$[1]$	9	15
$[1, 1]$	8	14
$[1, 1, 2, 3, 1, 1]$	4	7
$[1, 1, 2, 3, 1, 1, 2, 3, 1, 1]$	0	0

Therefore, for this p we have $R(t, u) = 1 + t^4u^7 + t^8u^{14} + t^9u^{15}$. Notice that if for some p no autocorrelation occurs, then we have $\mathcal{Q} = \{p\}$ and therefore $R(t, u) = 1$.

Finally, we define the *kernel* of a lattice path avoiding some pattern p as the following Laurent polynomial:

$$K(t, u) := (1 - tP(u))R(t, u) + t^{|p|}u^{h(p)}. \tag{2}$$

Also, in our case it can be shown that each root $u = u(t)$ of $K(t, u) = 0$ is either small (*i.e.*, $\lim_{t\to 0} u(t) = 0$) or large (*i.e.*, $\lim_{t\to 0} |u(t)| = +\infty$). The small roots are denoted by $u_1, \ldots, u_e$. We will also refer to them as the *small branches.*

Now we can state the enumeration results. Recall that t is the variable for the length of a path, and u is the variable for its final altitude.

▶ **Theorem 1.** *Let $\mathcal{S}$ be a set of steps, and let p be a pattern with steps from $\mathcal{S}$. Denote $\ell = |p|$, $b = h(p)$. Let $R(t, u)$ be the autocorrelation polynomial of the pattern p. Let $u_1, \ldots, u_e$ be the*

[2] A similar notion also appears in the work of Schützenberger on synchronizing words [31].

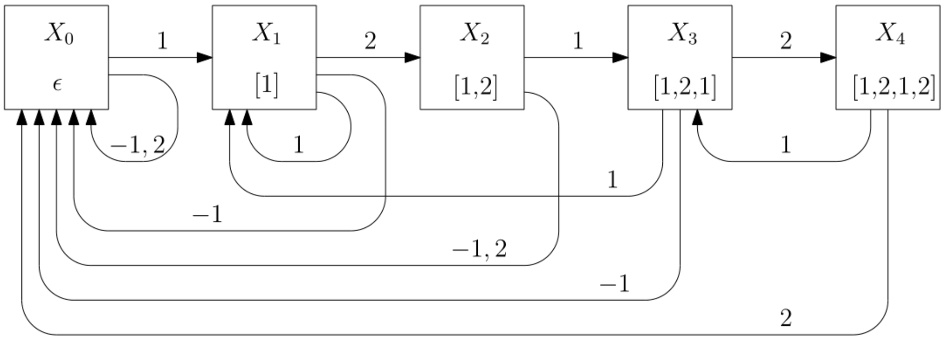


Figure 2 The automaton for the jumps $\mathcal{S} = \{-1, 1, 2\}$ and the pattern $p = [1, 2, 1, 2, -1]$. Any walk avoiding a given pattern p is associated with a similar automaton. It is in fact a *pushdown* automaton, in order to follow the positivity constraint. The matricial kernel method leads to the formulas of Theorem 1 for the corresponding generating functions, without having to solve a big algebraic system.

small roots of the kernel $K(t, u)$*, as defined in* (2)*. Then (under one additional constraint detailed in the proof), the generating functions of walks, bridges, meanders and excursions avoiding the pattern* p *are given by:*

$$W(t) = \frac{1}{1 - tP(1) + t^\ell / R(t,1)}, \tag{3}$$

$$B(t) = -\sum_{i=1}^{e} \frac{u_i'(t)}{u_i(t)} \frac{R(t, u_i)}{K_t(t, u_i)}, \tag{4}$$

$$M(t) = \frac{R(t,1)}{K(t,1)} \prod_{i=1}^{c} \left(1 - u_i(t)\right), \tag{5}$$

$$E(t) = \begin{cases} \frac{(-1)^{c+1}}{t} \prod_{i=1}^{c} u_i(t) & \text{if } b > -c, \\ \frac{(-1)^{c+1}}{t - t^\ell} \prod_{i=1}^{c} u_i(t) & \text{if } b = -c. \end{cases} \tag{6}$$

Proof. We refer to our companion article [1] for the proofs and the complete bivariate generating functions. The kernel $K(t, u)$ is in fact the determinant of $(\mathbf{I} - tA(u))^{-1}$, where $A(u)$ is the transition matrix encoding the stack automaton associated with the constrained walk (see Figure 2 below). The formulas then follow from an extension of the kernel method to matrix equations. (In fact, we presented above the simplified formulas for M and E, when p is what we call a *pseudomeander*, *i.e.* a lattice path which does not cross the x-axis, except, possibly, at its last step. If this is not the case, then we may have more than c small roots.) ◀

▶ Remark. Notice that for these four classes of lattice paths, forbidding a pattern of length 1 or using symbolic weights for the jumps recovers the formulas from Banderier and Flajolet [5].

4 Asymptotics of excursions avoiding a given pattern

The aim of this section is to characterize the asymptotics of the number of walks with jumps $\mathcal{S}$ avoiding a given pattern p.

▶ **Lemma 2** (Location of the dominant singularity)**.** *The dominant singularity (i.e. the nearest from zero) of $B(t)$ and $E(t)$ is ρ, the smallest real positive number where a small branch meets a large branch. (The branches refer to the solutions of $K(t,u)=0$, as defined in* (2)*).*

Proof. Lattice paths avoiding a given pattern can be generated by a pushdown automaton (see Figure 2). Accordingly, they can be generated by a context-free grammar, and their generating functions therefore satisfy a "positive" system of algebraic equations (see [15]). Therefore, the asympotic number of words of length n in such languages is of the form $C\rho^{-n}n^{\alpha}$. When the system is not strongly connected, α is either an integer (if ρ is a pole), either a dyadic number (if one has an iterated square root Puiseux singularity at ρ), as proven by Banderier and Drmota in [4]. For excursions, one has a strongly connected dependency graph (see Figure 2); the dominant singularity ρ (or, possibly, the dominant singularities) thus behaves like a square root. What is more, the cycle lemma (see the discussion on this in [5]) gives a correspondence between excursions and bridges, which implies that $E(t)$ and $B(t)$ have the same radius of convergence (this still holds when there is a forbidden pattern).

Now, because of the product formula (6) for excursions, one (or several) of the small branches have to follow this square root Puiseux behaviour. By Pringsheim's theorem, this has to be at a place $0<\rho\leq 1$; the geometry of the branches implies (see Table 2) that its location is where a large branch meets a small branch (because if the branching point comes from the intersection of small branches only, then their product will be regular). Therefore, ρ has to be the smallest real positive number where a small branch meets a large branch. ◀

▶ Remark. ρ is also the radius of convergence of meanders with negative or zero drift. For meanders with positive drift, the dominant pole of $1/K(t,1)$ will be the radius of convergence.

In order to avoid pathological cases, we now focus on *generic* walks.

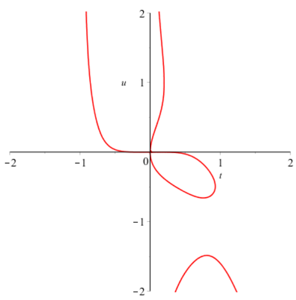
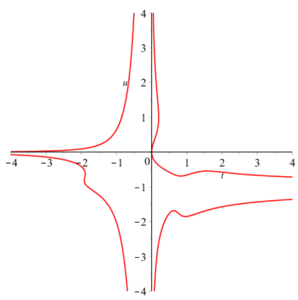
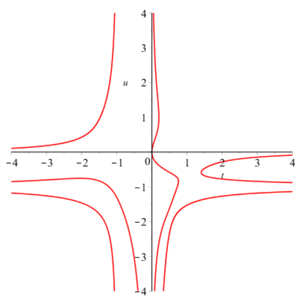
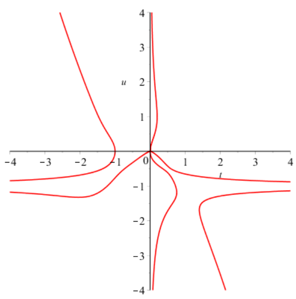
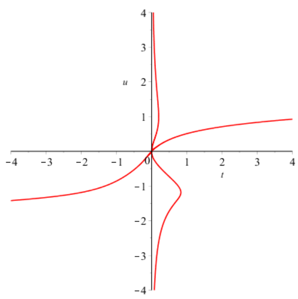
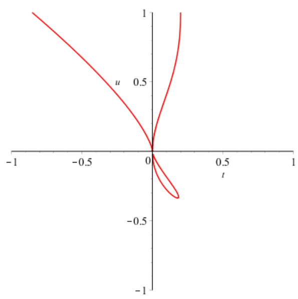
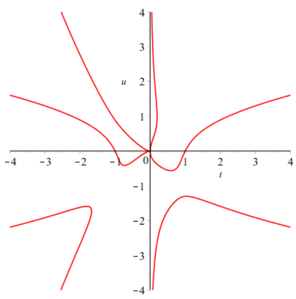
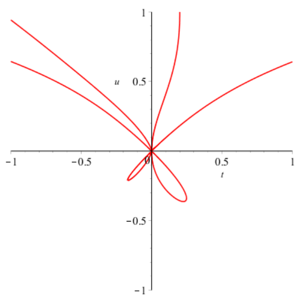
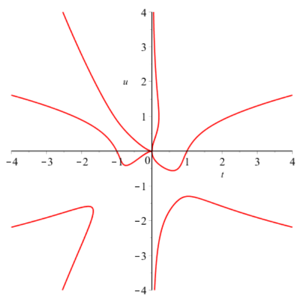
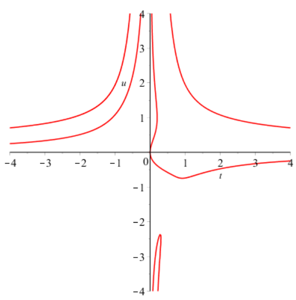
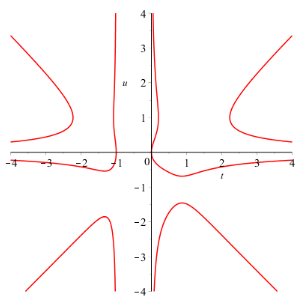
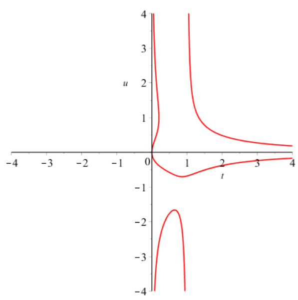
▶ **Definition 3** (Generic walks)**.** We call a constrained walk model "generic" if the following three properties hold.

- Property 1. The generating functions $B(t), M(t)$ and $E(t)$ are algebraic, not rational.
- Property 2. They have a unique domininant singularity.
- Property 3. No large negative branch (i.e. a branch such that $\lim_{z\to 0^+} u(z)=-\infty$) meets a small negative branch at ρ.

These three properties are very natural; we now comment more on them:

- For Property 1, it can be the case that the forbidden pattern leads to a degenerated model, in the sense that it is no more involving any stack and then we have words generated by a *regular* automaton (then, the generating functions are rational and the asymptotics are well understood). Example: $\mathcal{S}=\{-1,1\}$ and $p=[1,-1]$ or $p=[-1,-1]$.
- For Property 2, it is proven in [3] that multiple dominant singularities appear if and only if the gcd of the pairwise differences of the jumps is not 1. In this case, the asymptotics are obtained via [8, Theorem 8.8].
- For Property 3, we conjecture that it always holds. We have a proof for many classes of walks, but some remaining cases are tricky as it is possible to exhibit cases where one small negative branch meets a large negative branch, at some $\rho'>\rho$: This is e.g. the case for $\mathcal{S}=\{-2,-1,0,1,2\}$ and $p=[0,1,-2]$. Moreover, it is also possible that two small negative branches meet at ρ: This is e.g. the case for $\mathcal{S}=\{-2,1\}$ and $p=[1,-2,1,-2]$.

Table 2 Plot of the real branches of the kernel equation $K(t,u) = 0$, for several pattern p. This illustrates the diversity of behaviours. In all the examples, the set of jumps is $S = \{-2,-1,0,1,2\}$, and the pattern p is indicated. Note that due to a theorem of Pólya–Fatou–Carlson [14] on pure algebraic functions with integer coefficients (and therefore for generic walks), the first crossing between a small and large branch is at $0 < \rho < 1$ (*i.e.* $\rho = 1$ or any other root of $t - t^\ell$, cannot be the dominant singularity).

 $p = [1,-2,1,1,-2,1]$	 $p = [0,1,0,0,1,0]$	 $p = [0,0,1,2,0,0]$
 $p = [0,0,-1,-2,0,0]$	 $p = [-1,-2,-1,-2,-1,-2]$	 $p = [-1,-2,-1,-1,-2,-1]$
 $p = [-1,-1,0,-1,-1,0]$	 $p = [-2,-2,0,-2,-2,0]$	 $p = [-2,-1,1,-2,-1,1]$
 $p = [2,2,-1,2,2-1]$	 $p = [2,-1,-1,2,-1,-1]$	 $p = [2,-1,-1,2,-1,1]$

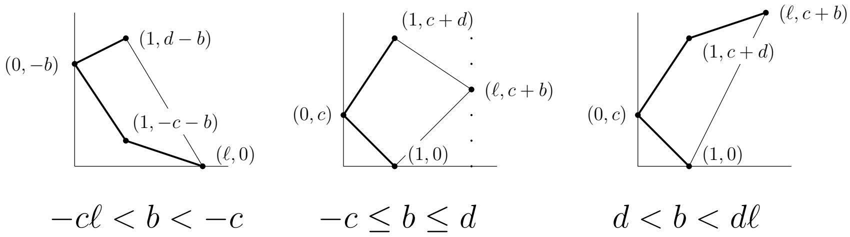


Figure 3 The three possibilities for the Newton polygon of the kernel $K(t, u)$ (the list is exhaustive if $R(t, u) = 1$). Each dot (i, j) corresponds to a monomial $t^i u^j$ of K. The slopes of the convex hull segments on the left give the Puiseux behaviour at 0 of the small and large roots u_i and v_j.

We observe that the behaviour of real branches of $K(t, u) = 0$ is much more complicated and diverse than that in the Banderier–Flajolet study. To recall, in their case there are always exactly two real *positive* branches (one small branch u_1 and one large branch v_1), and they meet at a singularity point $(t, u) = (\rho, \tau)$, where $u = \tau$ is the only positive number such that $P'(\tau) = 0$. In contrast, in our case we may have additional positive branches – even when the autocorrelation is trivial. Table 2 illustrates that we always have a small branch and one large branch whose shape in general resembles that of $u_1 \cup v_1$ from Banderier–Flajolet.

In one sense, the forbidden pattern gives a perturbation of the Banderier–Flajolet geometry of branches, and adds additional branches. A rigorous version of this intuition can be obtained by playing with a Boltzmann weight/Gibbs measure (like in statistical mechanics): moving the parameter v in a continuous way from 1 to 0 in the generating function $F(t, u, v)$ in the next section gives the explanation of these phenomena.

More information about these branches (on their Puiseux expansions) can be derived from the Newton polygon associated with the kernel (see [18] for a crisp presentation of the theory of Newton polygons for Puiseux expansion).

Equipped with all this information on the roots, and the way they cross, we can derive the following asymptotic results. Note that we use the notations $K_t(t, u)$ for $(\partial_t K)(t, u)$, and $K_{uu}(t, u)$ for $(\partial_u^2 K)(t, u)$. We start with asymptotics of walks on $\mathbb{Z}$ with a forbidden pattern.

▶ **Theorem 4** (Asymptotics of walks on $\mathbb{Z}$)**.** *Let ρ_K be the smallest positive root of $K(t, 1)$. For any generic model, the asymptotic number of walks of length n is:*

$$W_n \sim -\rho_K K_t(\rho_K, 1) R(\rho_K, 1) \rho_K^{-n}\,.$$

Proof. This follows from the partial fraction decomposition of $W(t) = \frac{R(t,1)}{K(t,1)}$. ◀

▶ **Theorem 5** (Asymptotics of excursions)**.** *Assume that we have a generic walk avoiding a pattern p which is a pseudomeander. Let $Y(t) := (-1)^{c-1} u_2(t) \cdots u_c(t)$. The number of excursions of size n satisfies*

$$E_n \sim Y(\rho) \sqrt{\frac{K_t(\rho, 1)}{2\pi\rho K_{uu}(\rho, 1)}} \cdot n^{-3/2} \rho^{-n}\,.$$

Proof. Since the walk is generic, $Y(t)$ is analytic for $|t| \leq \rho$. Thus the singular behaviour of $u_1(t)$ determines the singularity and the local behaviour of $E(t)$. We obtain:

$$E(t) \sim E(\rho) - Y(\rho) \sqrt{\frac{2K_t(\rho, \tau)}{\rho K_{uu}(\rho, \tau)}} \sqrt{1 - \frac{t}{\rho}}. \qquad ◀$$

▶ **Theorem 6** (Asymptotics of bridges). *Assume that we have a generic walk avoiding a pattern p. The number of bridges of size n satisfies*

$$B_n \sim -\frac{R(\rho,\tau)}{\tau K_t(\rho,\tau)}\sqrt{\frac{K_t(\rho,1)}{2\pi\rho K_{uu}(\rho,1)}}\cdot n^{-1/2}\rho^{-n}.$$

Proof. We know from Lemma 2 that $B(t)$ and $E(t)$ have the same radius of convergence. Thus, the singular behaviour of $u_1(t)$ determines the singularity and the local behaviour of $B(t)$. We have therefore

$$B(t) \sim -\frac{R(t,u_1(t))}{K_t(t,u_1(t))}\frac{u_1'(t)}{u_1(t)}$$

and plugging the singular expansion of u_1 into this formula yields the result. ◀

▶ **Theorem 7** (Asymptotics of meanders). *Let ρ_K be the smallest positive root of $K(t,1)$ (as in Theorem 4). Assume that the walk is generic and that p is a pseudomeander. Then the asymptotics of the coefficients of the meander generating function*

$$M(z) = (1-u_1(t))Y(t)R(t,1)/K(t,1) \qquad \textit{with } Y(t) := \prod_{i=2}^{c}(1-u_i(t))$$

is given by

$$M_n \sim R(\rho,1)Y(\rho)\sqrt{\frac{2}{\pi\rho K_t(\rho,1)K_{uu}(\rho,1)}}\cdot n^{-1/2}\rho^{-n} \qquad \textit{(for } \rho_K=\rho\textit{)},$$
$$M_n \sim -\frac{Y(\rho_K)R(\rho_K,1)}{\rho_K K_t(\rho_K,1)}\cdot \rho_K^{-n} \qquad \textit{(for } \rho_K<\rho\textit{)},$$
$$M_n \sim \frac{R(\rho,1)Y(\rho)}{K(\rho,1)}\sqrt{\frac{\rho K_t(\rho,1)}{2\pi K_{uu}(\rho,1)}}\cdot n^{-3/2}\rho^{-n} \qquad \textit{(for } \rho_K>\rho\textit{)}.$$

Proof. To prove the first assertion, observe that $\rho_K=\rho$ is equivalent to $\tau=1$. The dominant singularity of the generating function $M(t)=(1-u_1(t))Y(t)R(t,1)/K(t,1)$ is at $\rho_K=\rho$ and it originates from a simple zero in the denominator $K(t,u)$ and from u_1. The singular expansion from $u_1(t)$ at ρ gives (we use $\kappa(t) := -\frac{1}{tK_t(t,1)}$):

$$M(t) \sim R(\rho,1)Y(\rho)\kappa(\rho)\sqrt{\frac{2\rho K_t(\rho,1)}{K_{uu}(\rho,1)}}\left(1-\frac{t}{\rho}\right)^{-1/2} = \frac{R(\rho,1)Y(\rho)\sqrt{2}}{\sqrt{\rho K_t(\rho,1)K_{uu}(\rho,1)}}\left(1-\frac{t}{\rho}\right)^{-1/2}.$$

In the case $\rho_K<\rho$ we have $\tau\neq 1$ and thus $K(\rho,1)>0$. Hence the generating function has the dominant singularity ρ_K which comes from the kernel only. This implies

$$M(t) \sim Y(\rho_K)R(\rho_K,1)\kappa(\rho_K)\frac{1}{1-t/\rho_K}.$$

In the last case, $\rho_K>\rho$, u_1 has a square-root type singularity before $K(t,1)$ becomes singular. Singularity analysis thus gives the last claim of the theorem, via the following Puiseux expansion at the dominant singularity ρ

$$M(t) \sim M(\rho) + \frac{R(\rho,1)Y(\rho)}{K(\rho,1)}\sqrt{\frac{2\rho K_t(\rho,1)}{K_{uu}(\rho,1)}}\sqrt{1-\frac{t}{\rho}}.$$ ◀

Caveat: We are aware that several constants in these theorems can be further simplified, but we kept them like this in order to help the reader to follow the proofs (just sketched here, due to the page limit).

The theorems above for excursions and meanders are stated when the pattern p is a pseudomeander; there is a similar result for any pattern, but the proof goes through a wider disjunction of cases to handle, as then the closed form for the generating function is no more the same. We will handle this in the full version of this article.

These asymptotics also allow to get results on limit laws, as presented in the next section.

5 Limit law for the number of occurrences of a given pattern

Our approach also allows to count the number of occurrences of a pattern in paths. As usual, an occurrence of p in w is any substring of w that coincides with p, and when we count them we do not require that the occurrences will be disjoint. For example, the number of occurrences of 11 in 1111 is 3. We use the same notations than in Section 3. Then one has

▶ **Theorem 8** (Gaussian limit laws for occurrences). *Let X_n be the random variable which counts the number of occurrences of a pattern in a generic walk, bridge, meander, excursion model. Then X_n has a Gaussian limiting distribution with $\mathbb{E}[X_n] = \mu n + O(1)$ and $\mathbb{V}\mathrm{ar}[X_n] = \sigma^2 n + O(1)$ for some constants $\mu > 0$ and $\sigma^2 \geq 0$:*

$$\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n]\right) \to \mathcal{N}(0, \sigma^2).$$

Proof (sketch). The proof relies on the Gaussian limit laws for positive algebraic systems from [4, Theorem 9], which itself comes from following the dependency in the graph associated with the system, and applying Hwang's quasi-power theorem to each component. In this process, some positive variance conditions have to be checked on the formulas given by an equivalent of Theorem 1, with the additional variable v counting the number of occurrences of the pattern, and where the corresponding trivariate kernel is

$$K(t,u,v) := \det(I - tA) = (1-v)((1-tP(u))R(t,u) + t^\ell u^b) + v(1 - tP(u)). \tag{7}$$

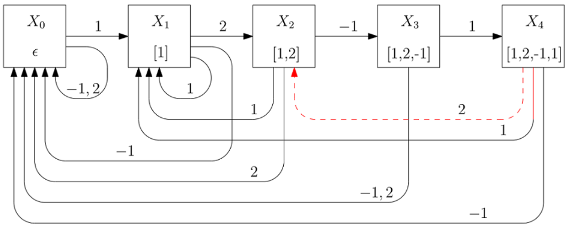
This comes from the associated automaton (as illustrated in Figure 4), and its adjacency matrix A. Note that for $v = 0$ we get the kernel from the avoidance case (see equation (2)), and for $v = 1$ we get $1 - tP$ (which is, as expected, the kernel from [5]).

To show the relation (7), we use a method adapted from [22, p. 60]. Let $W \equiv W(t,u,v)$ and $W_p \equiv W_p(t,u,v)$ be the generating functions of all words and words ending with p, respectively, where v counts the number of occurrences of p. We show the following two identities:

$$1 + WtP = W - W_p + v^{-1} W_p\,, \tag{8}$$

$$Wt^\ell u^b = v^{-1} W_p R - (R-1) W_p\,. \tag{9}$$

This system is readily solved and gives W as a rational function with denominator the right-hand side of (7). Since it is an irreducible polynomial, with degree ℓ in t, this denominator times a polynomial factor $Q(t,u,v)$ has to be equal to $\det(I - tA)$. In fact, $Q(t,u,v) = 1$. Indeed, an inspection of the degrees of the product shows that they cannot be higher than the degrees of the determinant of $I - tA$, and multiplying the denominator by a non constant Q would contradict this. Now, setting $v = 1$ gives that $1 - tP = \det(I - tA) = (1 - tP)Q(t,u,1)$ and thus $Q = 1$. This shows (7).



$$A = \begin{bmatrix} u^{-1}+u^2 & u & \cdot & \cdot & \cdot \\ u^{-1} & u & u^2 & \cdot & \cdot \\ u^2 & u & \cdot & u^{-1} & \cdot \\ u^{-1}+u^2 & \cdot & \cdot & \cdot & u \\ u^{-1} & u & u^2 v & \cdot & \cdot \end{bmatrix}$$

Figure 4 Pushdown automaton for the set of jumps $\mathcal{S} = \{-1, 1, 2\}$ and the pattern $p = [1, 2, -1, 1, 2]$. In dashed red we marked the arrow from the last state ($X_{\ell-1}$) labelled by the last letter of the pattern (a_ℓ). Marking this transition with v leads to formulas involving the kernel $K(t, u, v) = \det(I - tA)$ as given in Equation (7), where A is the adjacency matrix of this automaton.

To show (8), take a word and add a letter to it. If the resulting word does not end with p, it is counted by $W - W_p$; if it does, it is counted by $v^{-1}W_p$. To show (9), take a word w and add the pattern p to it. This creates a number $j \geq 1$ of new occurrences of p. The path wp can be written in j ways as $w'r$, where w' ends with a new occurence of p and r is an autocorrelation factor, or $j - 1$ ways if we impose that $r \neq \varepsilon$. It is therefore counted with a factor $v + \cdots + v^j$ by $W_p R$ and with a factor $v + \cdots + v^{j-1}$ by $(R - 1)W_p$, and the result follows. ◀

6 Conclusion

In this article, we presented a unifying way which gives the asymptotics of all families of lattice paths with a forbidden pattern, and we proved that the number of occurrences of a given pattern is normally distributed. The same approach would, for instance, allow to do the asymptotics of walks having exactly m occurrences of a given pattern, or to consider patterns which are no longer a word but a regular expression.

It is also nice that our approach gives a *method* (let us call it the *vectorial kernel method*) to solve in an efficient way the question of the enumeration and asymptotics of words generated by a pushdown automaton (or words belonging to the intersection of an algebraic language and a rational language). What is more, it is possible to use our functional equation approach to analyse the intersection of two algebraic languages. Note that testing if this intersection is empty is known to be an undecidable problem, even for deterministic context-free grammars (see e.g. [24]), so we cannot expect too much from a generic method in this case. However, we can specify a little bit more the type of system of functional equations we get: indeed this problem is related to automata with two stacks, which, in turn, are known to have the same power as a Turing machine; the evolution of these two stacks corresponds to lattice paths in the quarter plane (with steps of arbitrary length), the complexity of the problem is reflected by the fact that one can then get generating functions which are no more algebraic, D-finite, or differentially-algebraic, and we do not expect some universal nice results here, but a wider zoo of behaviours.

However, no doubt that all these cases will be new instances of what Flajolet and Sedgewick called *Borges's Theorem*: Any pattern which is not forbidden by design will appear a linear number of times in large enough structures, with Gaussian fluctuations.

For sure, it is more a metatheorem, a natural credo, so it is always worthwhile to establish this claim rigorously. Naturally, may it be with tools of probability theory or of analytic combinatorics, there is always some technical conditions to check to ensure this claim. In this article, our closed form expressions for the generating functions were one of the keys, together with the universal behaviour of the small branches. This allowed us to prove this Gaussian behaviour for the number of occurrences of any given pattern. Year after year, this claim is established for more and more combinatorial structures (it was done for patterns in Markov chains, trees, maps, permutations, context-free grammars, and now... lattice paths!).

Let us end with the passage of Flajolet and Sedgewick [22, p. 61] which explains where Borges's Theorem comes from:

> This property is sometimes called "Borges's Theorem" as a tribute to the famous Argentinian writer Jorge Luis Borges (1899-1986) who, in his essay *The Library of Babel*, describes a library so huge as to contain:
>
> > "Everything: the minutely detailed history of the future, the archangels' autobiographies, the faithful catalogues of the Library, thousands and thousands of false catalogues, the demonstration of the fallacy of those catalogues, the demonstration of the fallacy of the true catalogue, the Gnostic gospel of Basilides, the commentary on that gospel, the commentary on the commentary on that gospel, the true story of your death, the translation of every book in all languages, the interpolations of every book in all books."

References

1 Andrei Asinowski, Axel Bacher, Cyril Banderier, and Bernhard Gittenberger. Analytic combinatorics of lattice paths with forbidden patterns: enumerative aspects. In S. Klein, C. Martín-Vide, and D. Shapira, editors, *Language and Automata Theory and Applications. LATA 2018*, volume 10782 of *Lecture Notes in Computer Science*, pages 195–206. Springer, 2018. doi:10.1007/978-3-319-77313-1_15.

2 Axel Bacher and Mireille Bousquet-Mélou. Weakly directed self-avoiding walks. *Journal of Combinatorial Theory, Series A*, 118(8):2365–2391, 2011. doi:10.1016/j.jcta.2011.06.001.

3 Cyril Banderier. *Combinatoire analytique des chemins et des cartes.* Ph.D. thesis, Université Paris VI, 2001. URL: https://lipn.univ-paris13.fr/~banderier/Papers/these.pdf.

4 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.

5 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.

6 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. URL: https://hal.inria.fr/hal-01184683/document.

7 Cyril Banderier and Pierre Nicodème. Bounded discrete walks. *Discrete Math. Theor. Comput. Sci.*, AM:35–48, 2010. URL: https://dmtcs.episciences.org/2792/pdf.

8 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. Springer, 2018. To appear. URL: https://arxiv.org/abs/1606.08412.

9 Jean-Luc Baril and Armen Petrossian. Equivalence classes of Motzkin paths modulo a pattern of length at most two. *J. Integer Seq.*, 18(11):Article 15.7.1, 2015. URL: `https://cs.uwaterloo.ca/journals/JIS/VOL18/Baril/baril3.pdf`.

10 Antonio Bernini, Luca Ferrari, Renzo Pinzani, and Julian West. Pattern-avoiding Dyck paths. *Discrete Math. Theor. Comput. Sci. Proc.*, pages 683–694, 2013. URL: `https://dmtcs.episciences.org/2334/pdf`.

11 Mireille Bousquet-Mélou. Rational and algebraic series in combinatorial enumeration. In *International Congress of Mathematicians. Vol. III*, pages 789–826. EMS, 2006. URL: `https://hal.archives-ouvertes.fr/file/index/docid/277255/filename/icm.pdf`.

12 Mireille Bousquet-Mélou and Arnaud Jehanne. Polynomial equations with one catalytic variable, algebraic series and map enumeration. *Journal of Combinatorial Theory. Series B*, 96(5):623–672, 2006. `doi:10.1016/j.jctb.2005.12.003`.

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

14 Fritz Carlson. Über Potenzreihen mit ganzzahligen Koeffizienten. *Math. Z.*, 9(1-2):1–13, 1921. `doi:10.1007/BF01378331`.

15 Noam Chomsky and Marcel-Paul Schützenberger. The algebraic theory of context-free languages. In *Computer programming and formal systems*, pages 118–161. North-Holland, Amsterdam, 1963.

16 Nachum Dershowitz and Shmuel Zaks. More patterns in trees: up and down, young and old, odd and even. *SIAM J. Discrete Math.*, 23(1):447–465, 2008/09. `doi:10.1137/070687475`.

17 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`.

18 Jean Dieudonné. *Calcul infinitésimal.* Hermann, Paris, 2 edition, 1980. 1st ed. in 1968: 479pp., there is also an English translation of the 1st ed. in 1971, 427pp.

19 Yun Ding and Rosena R. X. Du. Counting humps in Motzkin paths. *Discrete Appl. Math.*, 160(1-2):187–191, 2012. `doi:10.1016/j.dam.2011.08.018`.

20 Philippe Duchon. On the enumeration and generation of generalized Dyck words. *Discrete Mathematics*, 225(1-3):121–135, 2000. `doi:10.1016/S0012-365X(00)00150-3`.

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 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`.

24 John E. Hopcroft, Raghavan Motwani, and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation.* Pearson, 3rd edition, 2006. (First edition by Addison-Wesley in 1979). URL: `http://infolab.stanford.edu/~ullman/ialc.html`.

25 Jacques Labelle and Yeong Nan Yeh. Generalized Dyck paths. *Discrete Mathematics*, 82(1):1–6, 1990. `doi:10.1016/0012-365X(90)90039-K`.

26 Kostas Manes, Aristidis Sapounakis, Ioannis Tasoulas, and Panagiotis Tsikouras. Strings of length 3 in Grand-Dyck paths and the Chung–Feller property. *Electron. J. Combin.*, 19(2):Paper 2, 10, 2012. URL: `http://www.combinatorics.org/ojs/index.php/eljc/article/view/v19i2p2/pdf`.

27 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`.

28 Donatella Merlini, D. G. Rogers, R. Sprugnoli, and M. Cecilia Verri. Underdiagonal lattice paths with unrestricted steps. *Discrete Appl. Math.*, 91(1-3):197–213, 1999. `doi:10.1016/S0166-218X(98)00126-7`.

29 Youngja Park and Seung Kyung Park. Enumeration of generalized lattice paths by string types, peaks, and ascents. *Discrete Math.*, 339(11):2652–2659, 2016. `doi:10.1016/j.disc.2016.04.024`.

30 Marcel-Paul Schützenberger. On context-free languages and push-down automata. *Information and Control*, 6:246–264, 1963.

31 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`.

32 Richard P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, 2nd edition, 2011.