Analytic combinatorics of lattice paths with forbidden patterns

Andrei Asinowski[•], Axel Bacher[*], Cyril Banderier[*], Bernhard Gittenbeger[•]




[•]Technische Universität Wien & [*]Université Paris Nord

Patterns in combinatorial structures

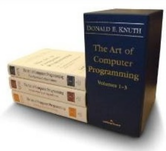
Enumerative and asymptotic aspects:

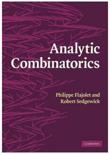
- patterns in permutations (links with sorting algorithms: Knuth)
- patterns in graphs, trees, maps (links with logics: Noy)
- patterns in words (generated by an automata / Markov Chain)
- patterns in integer compositions, partitions (links with number theory: Ramanujan)
- patterns in lattice paths (links with bioinformatics)

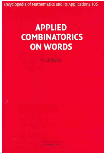
Typical questions :

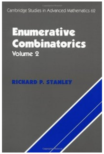
* What is number of structures of size n with k occurrences of the pattern?
* Is there a nice formula for the generating function?
* Asymptotic behaviour, limit laws?
* Generation of these objects?

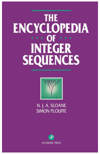
Patterns in words and combinatorial structures











Theorem (Chomsky-Schützenberger 1963)

Context free grammars (and pushdown automata) have algebraic generating functions: $F(z) = \sum f_n z^n \in Alg$.

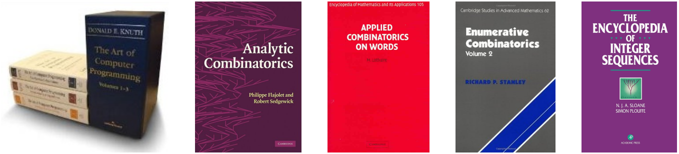
Now consider a language $\mathcal{L}$, in which we remove all words containing pattern p:

$\mathcal{L} \backslash (\Sigma^*.p.\Sigma^*)$ (therefore $\in \text{Alg} \cap \text{Rat} = \text{Alg}$, whenever $\mathcal{L}$ is algebraic).

- $\Rightarrow$ Good news: algebraic functions are well studied, lot of tools to deal with them (Gröbner bases, Newton polygon, singularity analysis).
- $\Rightarrow$ Bad news: even simple problem can lead to algebraic equation of degree more than the number of particules in the universe.

So what to do? Can we hope for "compact" formulas?

Patterns in words and combinatorial structures



Theorem (Chomsky-Schützenberger 1963)

Context free grammars (and pushdown automata) have algebraic generating functions: $F(z) = \sum f_n z^n \in Alg$.

Now consider a language $\mathcal{L}$, in which we remove all words containing pattern p:

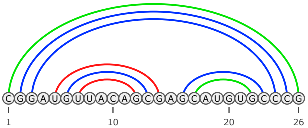
$$\mathcal{L} \backslash (\Sigma^*.p.\Sigma^*) \quad \text{(therefore} \quad \in \text{Alg} \cap \text{Rat} = \text{Alg, whenever } \mathcal{L} \text{ is algebraic).}$$

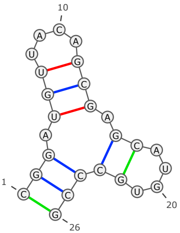
- $\Rightarrow$ Good news: algebraic functions are well studied, lot of tools to deal with them (Gröbner bases, Newton polygon, singularity analysis).
- $\Rightarrow$ Bad news: even simple problem can lead to algebraic equation of degree more than the number of particules in the universe.

So what to do? Can we hope for "compact" formulas?

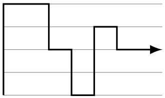
Yes, thanks to the vectorial kernel method and analytic combinatorics.

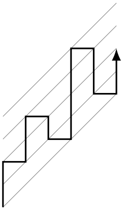
Patterns in lattice paths: applications

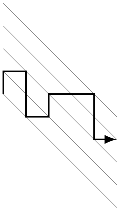




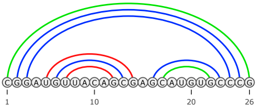
Patterns in lattice paths have applications to the study of secondary structure of RNA [see Yann Ponty's talk]

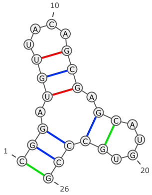


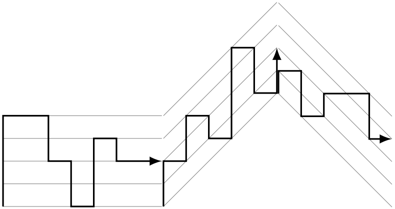




Some models of self-avoiding walks are encoded by partially directed lattice paths avoiding a pattern (see [Bacher–Bousquet-Melou 2011]).







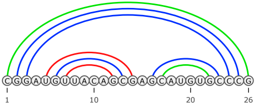
many articles on peaks/valleys/humps in Motzkin paths (jumps$= \{+1, 0, -1\}$) :

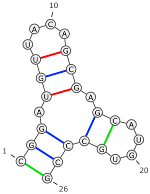
[Krattenthaler 2000, 2015] Surveys on Lattice path enumeration. Handbook of enumerative combinatorics (links with continued fractions, trees, permutations, determinants, ...)

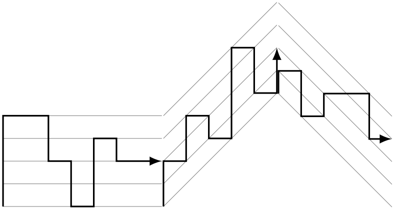
[Niederhausen; Eu; Yeh; Parvanien; Elizalde, Rubey; Park; Deutsch, Shapiro, Sulanke, Woan; Brennan, Prodinger; Ferrari, Rinaldi, Pergola, Pinzani; Bacher, Bousquet-Mélou; Ponty; Mansour, Shattuck; Sapounakis, Tasoulas, Tsiroukas; ...]

Is it possible to get a more unified point of view?

Patterns in lattice paths: vivid topics







many articles on peaks/valleys/humps in Motzkin paths (jumps$=\{+1,0,-1\}$) :

[Krattenthaler 2000, 2015] Surveys on Lattice path enumeration. Handbook of enumerative combinatorics (links with continued fractions, trees, permutations, determinants, ...)

[Niederhausen; Eu; Yeh; Parvanien; Elizalde, Rubey; Park; Deutsch, Shapiro, Sulanke, Woan; Brennan, Prodinger; Ferrari, Rinaldi, Pergola, Pinzani; Bacher, Bousquet-Mélou; Ponty; Mansour, Shattuck; Sapounakis, Tasoulas, Tsiroukas; ...]

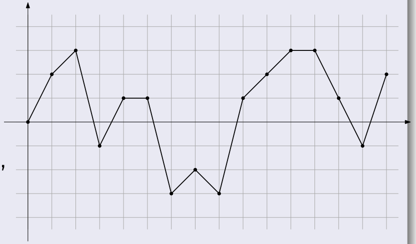
Is it possible to get a more unified point of view?

Yes, thanks to the vectorial kernel method and analytic combinatorics ;-)

Basic definitions

Definitions

Set of steps (S): a finite subset of $\mathbb{Z}$



Lattice path with steps from S:
A word $w = [v_1, v_2, \ldots, v_n]$ with v_i's $\in S$.
Visualized as the polygonal line obtained by appending vectors $(1, v_1), (1, v_2), \ldots, (1, v_n)$, starting at the origin.

Length: $|w| = n$,
Final altitude: $h(w) = v_1 + v_2 + \cdots + v_n$.

Notations: the kernel and its small roots

$$F(t,u) = \sum_{n\geq 0} \left(\sum_k f_{n,k} u^k \right) t^n$$

t is the variable for the length,
u is the variable for the final altitude.

$P(u)$ is the **steps polynomial** of S:

$$P(u) := \sum_{s \in S} u^s = u^{-c} + \ldots + u^d.$$

$K(t,u) := 1 - tP(u)$ is the **kernel**.

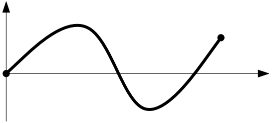
The kernel equation $K(t,u) = 0$, as an equation for u, has

- c many **small roots** ($\lim_{t\to 0} u(t) = 0$),
- and d many **large roots** ($\lim_{t\to 0} |u(t)| = \infty$).

The small roots are denoted by $u_1(t), \ldots, u_c(t)$.

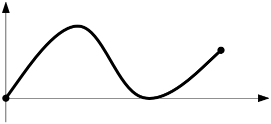
Enumeration and asymptotics of directed lattice paths

[Banderier and Flajolet 2002]



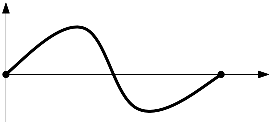
walks

$$W(t,u) = \frac{1}{1 - tP(u)}$$



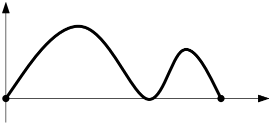
meanders

$$M(t,u) = \frac{1}{u^c(1 - tP(u))} \prod_{i=1}^{c} (u - u_i(t))$$



bridges

$$B(t) = t \sum_{i=1}^{c} \frac{u_i'(t)}{u_i(t)}$$



excursions

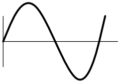
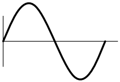
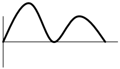
$$E(t) = \frac{(-1)^{c+1}}{t} \prod_{i=1}^{c} u_i(t)$$

Folklore example: Dyck paths $S = \{-1, 1\}$

jumps $P(u) = u^{-1} + u$

the kernel is $K(t,u) = 1 - t(u^{-1} + u)$

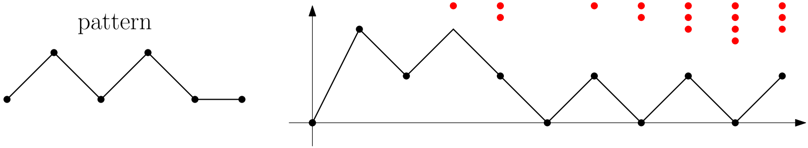
the small root is $u_1(t) = \dfrac{1 - \sqrt{1 - 4t^2}}{2t}$

	$W(t,u) = \dfrac{1}{1 - tP(u)}$	$W(t,u) = \dfrac{1}{1 - 2t}$
	$B(t) = t \sum_{i=1}^{c} \dfrac{u_i'(t)}{u_i(t)}$	$B(t) = \dfrac{1}{\sqrt{1 - 4t^2}}$
	$E(t) = \dfrac{(-1)^{c+1}}{t} \prod_{i=1}^{c} u_i(t)$	$E(t) = \dfrac{1 - \sqrt{1 - 4t^2}}{2t^2}$

Our work: Lattice paths with forbidden patterns

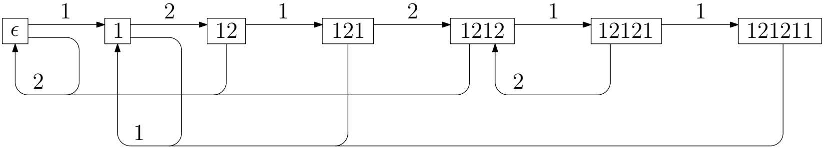
We fix a **pattern** – a word $p = [a_1, \ldots, a_\ell]$,
and we want to enumerate lattice paths that avoid p,
that is, do not contain p as a substring.

To each proper prefix of p we assign a *state*:
$X_i = [a_1, \ldots, a_i]$, where $0 \leq i \leq \ell - 1$. We say that a path p is in the state X_i after a certain step if by this point it "accumulated" $[a_1, \ldots, a_i]$ as its suffix:



Associated Automaton

This gives raise to a finite automaton. Example: $p = [1, 2, 1, 2, 1, 1, 2]$

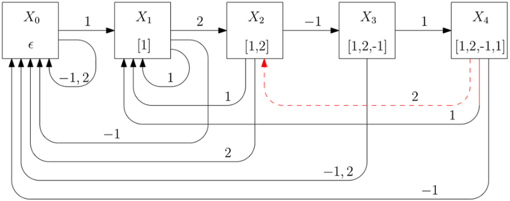


To keep track of altitude, for each jump s, one multiplies the corresponding transitions by u^s (it mimicks an additional stack, like a pushdown automata). This gives an adjacency matrix $A(u)$.
To each state X_i is associated a generating function $W_i(t, u)$ of walks that terminate in this state.
Thus, the generating function of walks avoiding the pattern p is

$$W(t, u) = W_1(t, u) + W_2(t, u) + \cdots + W_\ell(t, u)$$

Challenge: can we give a more explicit expression for W? How this approach can work if one needs to encode non-negativity constraints (e.g. for excursions $E(t)$ and meanders $M(t, u)$)?

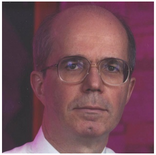
Automaton to count occurrences of the pattern



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

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

Autocorrelation polynomial



[Guibas–Odlyzko 1980]

A **presuffix** of a word p is a non-empty string that appears in p both as a prefix and as a suffix. The **autocorrelation polynomial** of the word p is

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

where $\mathcal{Q}$ is the set of all the presuffixes of p (and bar denotes complement to p).

Example: For $p = [1,1,0,1,1,0,1,1]$, presuffix(p)$=$ $\{$1, 11, 11011,p $\}$ so we have $R(t,u) = 1 + t^3u^2 + t^6u^4 + t^7u^5$.

If p has no proper presuffixes, then we have $R(t,u) = 1$ (*trivial autocorrelation*).

The kernel theorem

Theorem

Let A be the transfer matrix of the automaton that was described above. Then we have:

1. $K(t,u) := |I - tA| = (1 - tP(u))R(t,u) + t^{|p|}u^{h(p)}$.
2. $\vec{1_0}\ \mathrm{adj}(I - tA)\ \mathbb{1} = R(t,u)$,
   where $\vec{1_0} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$ *and* $\mathbb{1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^\top$.

The kernel theorem

Theorem

Let A be the transfer matrix of the automaton that was described above. Then we have:

1. $K(t,u) := |I - tA| = (1 - tP(u))R(t,u) + t^{|p|}u^{h(p)}$.
2. $\vec{1_0}\ \mathrm{adj}(I - tA)\ \mathbb{1} = R(t,u)$,
   where $\vec{1_0} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$ *and* $\mathbb{1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^\top$.

In particular, in the case of trivial autocorrelation we have
$K(t,u) = |I - tA| = 1 - tP(u) + t^{|p|}u^{h(p)}$ *and* $\vec{1_0}\ \mathrm{adj}(I - tA)\ \mathbb{1} = 1$.

Summary of results (BF vs. ABBG)

	[Ban-Fla 2002]	[Asi-Bac-Ban-Git 2018]
$W(t,u)$	$\dfrac{1}{K(t,u)}$	$\dfrac{R(t,u)}{K(t,u)}$
$B(t)$	$t\sum_{i=1}^{c}\dfrac{u_i'(t)}{u_i(t)}$	$-t\sum_{i=1}^{e}\dfrac{u_i'(t)}{u_i(t)}\dfrac{R(t,u_i)}{K_t(t,u_i)}$
$M(t,u)$	$\dfrac{1}{u^c K(t,u)}\prod_{i=1}^{c}(u-u_i(t))$	$\dfrac{R(t,u)}{u^c K(t,u)}\prod_{i=1}^{c}(u-u_i(t)) \quad (*)$
$E(t)$	$\dfrac{(-1)^{c+1}}{t}\prod_{i=1}^{c}u_i(t)$	$\dfrac{(-1)^{c+1}}{t}\prod_{i=1}^{c}u_i(t) \qquad (*)$

Generating function of meanders (proof, part 1/2)

We have

$$(M_1 \ \cdots \ M_\ell) = \mathbf{J} + t\,(M_1 \ \cdots \ M_\ell)\,A - t\,[u^{<0}]\,((M_1 \ \cdots \ M_\ell)\,A),$$

$$(M_1 \ \cdots \ M_\ell) = \vec{1_0} + t\,(M_1 \ \cdots \ M_\ell)\,A - \tfrac{t}{u^c} F(t,u)\vec{1_0}$$

$$(M_1 \ \cdots \ M_\ell)(I - tA) = \left(1 - \frac{t}{u^c} F(t,u)\right)\mathbf{J}$$

$$(M_1 \ \cdots \ M_\ell) = \left(1 - \frac{t}{u^c} F(t,u)\right)\mathbf{J}\,\frac{\mathrm{adj}(I - tA)}{|I - tA|}$$

$$M(t,u) = \left(1 - \frac{t}{u^c} F(t,u)\right)\frac{R(t,u)}{K(t,u)}$$

Generating function of meanders (proof, part 1/2)

$$(M_1 \ \cdots \ M_\ell)(I - tA) = \left(1 - \frac{t}{u^c} F(t,u)\right)\mathbf{J} \qquad (*)$$

The kernel, $K(t,u) = |I - tA| = (1 - t(u^{-c} + \cdots + u^d))R(t,u) + t^\ell u^h$, has c small roots: $u_1(t), \ldots, u_c(t)$ (because $-c < 0 \leq h$).

Let $i \in \{1, 2, \ldots, c\}$. Substitute $u = u_i(t)$ into $(*)$. The matrix $(I - tA)|_{u=u_i(t)}$ is singular. Let $\mathbf{v}$ be the eigenvector of this matrix with $\lambda = 0$ (its first component can be assumed to be 1). We multiply $(*)$ with $u = u_i(t)$ by $\mathbf{v}$ from the right and obtain $0 = u_i^c(t) - tF(t, u_i(t))$.

Thus, $u_1(t), u_2(t), \ldots, u_c(t)$ are roots of $u^c - tF(t,u) = 0$, a **polynomial** equation (in u) of degree c. Therefore, we have

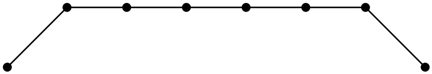
$$u^c - tF(t,u) = (u - u_1(t))(u - u_2(t)) \ldots (u - u_c(t)).$$

We obtain the meanders generating function

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

Example 1: Humps in Motzkin paths

$S = \{-1, 0, 1\}$, $p = [1, 0, 0, \ldots, 0, -1]$.



$$K(t, u) = 1 - t(u^{-1} + 1 + u) + t^{\ell},$$

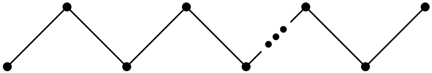
$$u_1(t) = \frac{1 - t + t^{\ell} - \sqrt{(1 + t + t^{\ell})(1 - 3t + t^{\ell})}}{2t}$$

$$B(t) = \frac{1}{\sqrt{(1 + t + t^{\ell})(1 - 3t + t^{\ell})}}$$

$\ell = 2$: OEIS A051286. "Whitney number of level n of the lattice of the ideals of the fence of order $2n$. Apparently the number of peakless grand Motzkin paths of length n."

Example 2: no zigzag Dyck paths

$S = \{-1, 1\}$, $p = [1, -1, 1, -1, \ldots, 1]$.



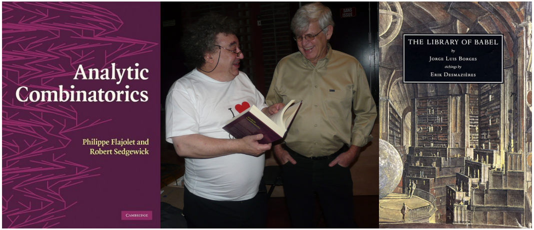
$$K(t, u) = 1 - t(u^{-1} + u)\frac{1 - t^{\ell+1}}{1 - t^2} + t^{\ell} u,$$

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

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

$$= 1 + t^2 + t^4 + 2\,t^6 + 4\,t^8 + 9\,t^{10} + 21\,t^{12} + 51\,t^{14} + \ldots$$

Flajolet & Sedgewick "Borges's theorem"

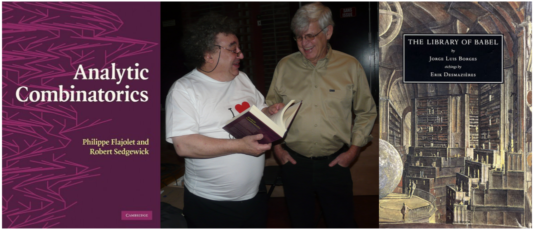


Jorge Luis Borges (1899-1986)
The library of Babel (1941).

The novel 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."

Flajolet & Sedgewick "Borges's theorem"



Jorge Luis Borges (1899-1986)
The library of Babel (1941).

"In any combinatorial structure, any pattern not forbidden by design, will appear linear number of time, with a Gaussian limit law."

Many examples:

- [Flajolet, Drmota, Soria, Hwang, Bender, Canfield 1993++] : OK if 1 algebraic equation, or 1 linear system
- [Gao–Wormald 2004] patterns in maps
- [Flajolet–Salvy–Nicodème 2002]: patterns in rational languages
- [Asinowski–Bacher–Banderier–Gittenberger 2018]: patterns in algebraic languages

Conclusion

- unified approach for counting lattice paths with a forbidden pattern
- unified approach for counting # occurrences of a pattern in lattice paths
- explicit expressions in terms of the autocorrelation polynomial
- extension of the kernel method to systems of equations
- can be applied to any pushdown automaton
- can be applied to intersection of context-free grammar and automaton
- analytic combinatorics approach: universality for the asymptotics, Borges's theorem: Gaussian limit law