Alcom-FT Technical Report Series
ALCOMFT-TR-01-8

ANALYTIC COMBINATORICS

Philippe Flajolet[1] and Robert Sedgewick[2]
[1] Algorithms Project, INRIA Rocquencourt, F-7850 Le Chesnay (France)
[2] Department of Computer Science, Princeton University, Princeton, NJ 08544 (USA)

Foreword

This report is part of a series whose aim is to present in a synthetic way the major methods and models in analytic combinatorics. The whole series, after suitable editing, is destined to be transformed into a book with the title

"Analytic Combinatorics"

The present report is

— Chapter 8, *"Functional Equations—Rational and Algebraic Functions"*,

and it is ge s INRIA Research Report RR 4103, 100 pages, January 2001. It is part of the following collection of Research Reports from INRIA:

— Chapters 1–3, "Counting and Generating Functions", RR 1888, 116 pages, 1993;
— Chapters 4–5, "Complex Asymptotics and Generating Functions", RR 2026, 100 pages, 1993;
— Chapter 6, "Saddle Point Asymptotics", RR 2376, 55 pages, 1994;
— Chapter 7, "Mellin Transform Asymptotics", RR 2956, 93 pages, 1996.
— Chapter 9, "Multivariate Asymptotics and Limit Distributions", RR 3162, 123 pages, May 1997.

Here, we detail the case of rational and algebraic functions and discuss systematically closure properties, the location of singularities, and consequences regarding combinatorial enumeration. The theory is applied to regular and context-free languages, finite state models, paths in graphs, locally constrained permutations, lattice paths and walks, trees, and planar maps.

Acknowledgements. This work was supported in part by the IST Programme of the European Union under contract number IST-1999-14186 (ALCOM-FT). Special thanks are due to Bruno Salvy for sharing his thoughts on the subject of algebraic asymptotics and to Pierre Nicodème for a thorough scan of an early version of the manuscript.

CHAPTER 8

Functional Equations—Rational and Algebraic Functions

Mathematics is infinitely wide, while the language that describes it is finite. It follows from the pigeonhole principle that there exist distinct concepts that are referred to by the same name. Mathematics is also infinitely deep and sometimes entirely different concepts turn out to be intimately and profoundly related.

— *Doron Zeilberger* [**101**]

I wish to God these calculations had been executed by steam.

— *Charles Babbage (1792-1871)*

This part of the book deals with classes of generating functions implicitly defined by linear, polynomial, or differential relations, globally referred to as *functional equations*. Functional equations arise in well defined combinatorial contexts and they lead systematically to well-defined classes of functions. One then has available a whole arsenal of methods and algorithmic procedures in order to simplify equations, locate singularities, and eventually determine asymptotics of coefficients. The corresponding classes are associated with algebraic closure properties together with a strong form of analytic "regularity" that constrains the location and nature of singularities. We shall detail here the case of rational functions (that arise from functional equations that are linear and from associated finite-state models) and algebraic functions (that arise from polynomial systems and "context-free" decompositions). A companion chapter will treat holonomic functions (defined by linear differential equations with coefficients themselves polynomial or rational functions that arise from a diversity of contexts including order statistics).

Rational functions come first in order of simplicity. Linear systems of equations occur systematically in all combinatorial problems that are associated with finite state models (themselves closely related to Markov chains), like paths in graphs, regular languages and finite automata, patterns in strings, and transfer matrix models of statistical physics. Singularities are by nature always *poles*. Consequently, the asymptotic analysis of coefficients of rational functions is normally achieved via localization of poles. Although difficulties may arise either in noncombinatorial contexts (nonpositive problems) or when dealing simultaneously with an infinite collection of functions (for instance, the generating functions of trees of bounded height or width), the situation is however often tractable: for positive systems, general theorems derived from the classical Perron-Frobenius theory of nonnegative matrices guarantee simplicity of the dominant positive pole.

Next in order of difficulty, there come the *algebraic functions* defined as solutions to polynomial equations. Such functions occur in connection with the simplest nonlinear combinatorial models, namely the class of context-free models. Such models cover many types of combinatorial trees (that are related to the theory of branching processes in probability theory) or walks (that are close to the probabilistic theory of random walks).

Algebraic properties include the possibility of performing elimination by devices like resultants or Groebner basis: in this way, a system of equations can always be reduced to a single equation. Singularities are in all cases of a simple form—they are *branch points* corresponding locally to *fractional power series*, also known as Newton–Puiseux expansions. Accordingly, the asymptotic *shape* of the coefficients of any given algebraic function is then entirely predictable by singularity analysis. However, nonlinear algebraic equations admit several solutions and a so-called "*connection problem*" has to be solved by consideration of the global geometry of the algebraic curve defined by a polynomial equation. Again, fortunately, the situation somewhat simplifies for positive algebraic systems that capture most of combinatorial applications.

Last but not least there come the *holonomic functions* that encompass rational and algebraic functions. They are defined as solutions of linear differential equations with rational function coefficients and they arise in a number of contexts, from order statistics to regular graphs. It has been realized over the last two decades, by Stanley, Lipshitz, and Zeilberger most notably, that these functions enjoy an extremely rich set of nontrivial closure properties. In particular, they encapsulate most of what is known to admit of closed form in combinatorial analysis. Algebraically, holonomic functions are objects that can be specified by a finite amount of information so that the identities they satisfy is *decidable*. For instance, even a restrictive consequence of this theory leads to an interesting fact summarized by Zeilberger's aphorism "*All binomial identities are trivial*". Analytically, the nature of singularities is again governed by simple laws, a fact that derives from turn-of-the-twentieth-century studies of singularities of linear differential equations (by Schwartz, Fuchs, Birkhoff, Poincaré, and others). The classification involves a fundamental dichotomy between what is known as *regular singularity* and *irregular singularity*. From there, like in the algebraic case, the asymptotic *shape* of the coefficients of any given holonomic function is predictable by singularity analysis (regular singularity) or saddle point analysis (irregular singularity). However, unlike in the algebraic case, the *connection problem* is not known to be decidable and a priori quantitative bounds (based on combinatorial reasoning) must sometimes be resorted to in order to prune singularities and completely solve an asymptotic question that arises from a combinatorial problem expressed by a holonomic generating function.

Finally, we make a brief mention here of equations of the composition type. There, strong closure properties tend to fade away. However, a nice set of techniques that have been put to use successfully in the analysis of some important combinatorial problems. For instance, the counting of balanced trees [**72**] and the distributional analysis of height in simple families of trees [**41, 42**] relate to analytic iteration theory. On another register, metric properties of general number representation systems and information sources [**97**] together with the corresponding analyses of digital trees [**24**] can be approached successfully by means of functional analytic methods, especially transfer operators.

The rational, algebraic, and holonomic classes

Exact combinatorial enumeration is best expressed in terms of generating functions, the recurrent theme of this book. In this chapter and its companion ("Functional Equations—Holonomic Functions"), three major classes are identified: the rational class, the algebraic class, and the holonomic class. Each class is a world of its own with specific algebraic properties and specific analytic properties.

At the level of algebra, everything is expressed in term of formal power series, since no consideration of convergence enters the discussion *a priori*. Given a domain $\mathbb{K}$, what is denoted by $\mathbb{K}[[z]]$ is the set of formal power series in the indeterminate z, which means the collection of formal sums,

$$f(z) = \sum_{n=0}^{\infty} f_n z^n, \qquad f_n \in \mathbb{K}.$$

We shall normally take $\mathbb{K}$ to be a field like the field $\mathbb{C}$ of complex numbers or the subfield $\mathbb{Q}$ of rational numbers. (We occasionally speak of $\mathbb{N}[[z]]$ or $\mathbb{Z}[[z]]$ but these will be regarded as simply denoting particular elements of $\mathbb{Q}[[z]]$ or $\mathbb{C}[[z]]$.) We then have, in order of increasing structural complexity and richness, three major classes of objects,

$$\mathbb{K}^{\mathrm{rat}}[[z]] \subset \mathbb{K}^{\mathrm{alg}}[[z]] \subset \mathbb{K}^{\mathrm{hol}}[[z]] \subset \mathbb{K}[[z]],$$

corresponding to the rational, algebraic, and holonomic subsets of $\mathbb{K}[[z]]$.

— *Rational series* denoted by $\mathbb{K}^{\mathrm{rat}}[[z]]$ are defined as solution of linear equations,

$$y \in \mathbb{K}^{\mathrm{rat}}[[z]] \text{ iff } y \in \mathbb{K}[[z]] \text{ and } a_1(z)y + a_0(z) = 0, \tag{1}$$

for some polynomials $a_0, a_1 \in \mathbb{K}[z]$.

— *Algebraic series* are defined as solutions of polynomial equations,

$$y \in \mathbb{K}^{\mathrm{alg}}[[z]] \text{ iff } y \in \mathbb{K}[[z]] \text{ and } \sum_{j=0}^{e} a_j(z) y^j = 0, \tag{2}$$

for a family of polynomials $a_j \in \mathbb{K}[z]$.

— *Holonomic series* are defined as solutions of linear differential equations,

$$y \in \mathbb{K}^{\mathrm{hol}}[[z]] \text{ iff } y \in \mathbb{K}[[z]] \text{ and } \sum_{j=0}^{e} a_j(z) \frac{d^j}{dz^j} y = 0, \tag{3}$$

for a family of polynomials $a_j \in \mathbb{K}[z]$.

It often proves convenient to extend rings into fields. The domain $\mathbb{K}(z)$ is the quotient field of the ring $\mathbb{K}[z]$, that is the set of fractions a/b, with $a, b \in \mathbb{K}[z]$ and it is a simple exercise to check that the definitions of (1), (2), (3) could have been phrased in an equivalent way by imposing that coefficients lie in $K(z)$ instead of $\mathbb{K}[z]$. The quotient field of $\mathbb{K}[[z]]$ is denoted by $\mathbb{K}((z))$ and is called the field of formal Laurent series. From its definition, a Laurent series contains a finite number of negative powers of z, and a formalist might enjoy an "identity" like $\mathbb{K}((z)) = \mathbb{K}[[z]][1/z]$. Then, in analogy to (1), (2), (3), one can define three subsets of $\mathbb{K}((z))$, namely, $\mathbb{K}^{\mathrm{rat}}((z)), \mathbb{K}^{\mathrm{alg}}((z)), \mathbb{K}^{\mathrm{hol}}((z))$ by looking at solutions in $\mathbb{K}((z))$ instead of $\mathbb{K}[[z]]$.

The definitions we have adopted are by means of a single equation. As it turn out, every class can be alternatively defined in terms of *systems of equations*. Theory grants us the fact that systems are reducible to single equations, in each of the three cases under consideration. The reduction can be seen as an *elimination* property: given a system Σ that defines simultaneously a vector $(y_1, \ldots, y_m)$ of solutions, each component, y_1 say, is definable by a single equation. For linear systems, this fact is granted by Gaussian elimination and by the theory of determinants ('Cramer's rule'). For polynomial systems, one may appeal either to Groebner basis elimination—an algorithmic process reminiscent of Gaussian elimination— or to resultants that are related to determinants. For differential systems, one may either appeal to an extension of Groebner bases to differential operators or to a method known under the name of "cyclic vectors"; see [**23**].

Each class carries with it a set of closure properties, some more obvious than others. Apart from the usual arithmetic operations, there is also interest in closure under Hadamard

1. Basic objects.

Domain	Notation	Typical element	($\mathbb{K}$ is a field)
Polynomials	$\mathbb{K}[z]$	$\sum_{j=0}^{e} c_j z^j$, with $c_j \in \mathbb{K}$	$\mathbb{K}[z]$ is a ring
Rational fractions	$\mathbb{K}(z)$	$\frac{A}{B}$ with $A, B \in \mathbb{K}[z]$	$\mathbb{K}(z)$ is a field
Formal power series	$\mathbb{K}[[z]]$	$\sum_{j=0}^{\infty} c_j z^j$	$\mathbb{K}[[z]]$ is a ring
Formal Laurent series	$\mathbb{K}((z))$	$\sum_{j=-e}^{\infty} c_j z^j$	$\mathbb{K}((z))$ is a field

2. Special classes. (Coefficients a_j may be taken in either $\mathbb{K}[z]$ or in $\mathbb{K}(z)$.)

Rational series	$\mathbb{K}^{\mathrm{rat}}[[z]]$	solution of $a_1(z)y + a_0(z) = 0$
Algebraic series	$\mathbb{K}^{\mathrm{alg}}[[z]]$	solution of $a_e(z)y^e + \cdots + a_0(z) = 0$
Holonomic series	$\mathbb{K}^{\mathrm{hol}}[[z]]$	solution of $a_e(z)\partial^e y + \cdots + a_0(z) = 0$ ($\partial \equiv \frac{d}{dz}$)

3. Closure properties.

Class	Elimination (System → Eq.)	$\pm$	$\times$	$\div$	$\odot$	∂	$\int$
Rat	Gaussian elim.; determinants	Y	Y	Y	Y	Y	N
Alg	Groebner bases; resultants	Y	Y	Y	N	Y	N
Hol	diff. Groebner bases; cyclic vector	Y	Y	N	Y	Y	Y

4. Singularities and coefficient asymptotics (simplified forms).

	Singularity		Coefficient form
Rat	$(z-\zeta)^{-m}$	(m integer)	$n^{m-1}\zeta^{-n}$
Alg	$(z-\zeta)^{-\alpha}$	($\alpha = \frac{p}{q} \in \mathbb{Q}$)	$n^{\alpha-1}\zeta^{-n}$
Hol [regular sing.]	$(z-\zeta)^{-\beta}(\log(z-\zeta))^k$	(β algebraic)	$n^{\beta-1}(\log n)^k\zeta^{-n}$

FIGURE 1. A summary of the definitions and major properties of the three classes of functions: Rational (Rat), Algebraic (Alg), and Holonomic (Hol).

products, that is, termwise product of series. Differentiation and integration of formal power series are defined in the usual way,

$$\partial\left(\sum_n f_n z^n\right) = \sum_n n f_n z^{n-1}, \qquad \int\left(\sum_n f_n z^n\right) = \sum_n \frac{f_n}{n+1} z^{n+1}.$$

The main closure properties of each class are summarized in Figure 1.

Next comes analysis. A major theme of this book is that asymptotic forms of coefficients are dictated by the expansions of functions at singularities. In fact functions associated to series that belong to any of the major three classes under consideration have a finite number of singularities and the nature of singularities is predictable.

Rational functions can only have poles for which the theory of Chapter 4 applies directly. Algebraic functions enjoy a richer singular structure, where fractional exponents occur: the corresponding expansions known since Newton are called Newton–Puiseux expansions and the conditions of singularity analysis are automatically satisfied. However, the inherently multivalued character of algebraic functions poses a specific "connection problem" as one needs to select the particular branch that is associated to any given combinatorial problem. Holonomic functions have well-classified singularities that fall into two categories called regular and irregular (also diversely known as first kind and second kind, or Fuchsian and non-Fuchsian). The simplest type, the regular type, introduces elements with exponents that may be arbitrary algebraic numbers together with integral powers of a logarithm. Again, locally the conditions of singularity analysis are automatically satisfied, but again a connection problem arises—how do coefficients in the expansion at the singularity relate to the initial conditions given at the origin? Typical elements of coefficients asymptotics that are encountered in this book are

$$(\frac{1+\sqrt{5}}{2})^n, \quad \frac{4^n}{\sqrt{\pi n^3}}, \qquad n^{(\sqrt{17}-3)/2},$$

and they reflect the nature of singularities present in each case. In effect, the first element is typical of rational asymptotics (here, a simple pole) and it arises in monomer-dimer tilings of the interval. The second one corresponds to the asymptotic form of Catalan numbers and its origin is a singular element of an algebraic function with exponent $1/2$. The third one corresponds to search cost in quadtrees and the algebraic number present in the exponent is indicative of a holonomic element.

Finally, positive functions, that is, solutions of positive systems, are the ones most likely to show up in elementary combinatorial applications. They are structurally more constrained and this fact is reflected to some extent by specific singularity and coefficient asymptotics. For instance, Perron-Frobenius theory says that, under certain natural conditions of "irreducibility", a unique dominant pole (that is simple) occurs in rational asymptotics. Similar conditions on positive algebraic systems constrain the singular exponent to be equal to $\frac{1}{2}$, hence the characteristic factor $n^{-3/2}$ present in the asymptotic form of so many enumerative problems.

In this report, we consider the rational and algebraic classes in turn. First, at an algebraic level, we state an elimination theorem and establish major closure properties. Next comes analysis, with its batch of singular asymptotics and matching coefficient forms. Applications (including regular and context-free specifications) illustrate the general theory in some important combinatorial situations.

1. Algebra of rational functions

Rational functions are the simplest of all objects considered in this chapter. They are naturally defined as quotients of polynomials (Definition 8.1) or alternatively as components of solutions of linear systems (Theorem 8.1) a form that is especially convenient for combinatorial enumeration. Coefficients of rational functions satisfy linear recurrences with constant coefficients and they also admit an explicit form, called "exponential polynomial" (Theorem 8.1) that is directly related to the location and multiplicity of poles and to the corresponding asymptotic behaviour of coefficients (Theorem 8.4). Closure properties are stated as Theorem 8.2 while easy combinatorial forms for coefficients of rational functions are given in Theorem 8.3 The asymptotic side of rational functions is the subject of Section 2. An important class of positive systems has poles whose location and nature can be predicted using an extension of the Perron-Frobenius theory of positive matrices (Theorems 8.5 and 8.6).

DEFINITION 8.1. *A power series $f(z) \in \mathbb{C}[[z]]$ is said to be rational if there exist two polynomials $P(z), Q(z)$, with $Q(0) \neq 0$, such that $Q(z)f(z) - P(z) = 0$, i.e.,*

$$f(z) = \frac{P(z)}{Q(z)}.$$

Clearly, we may always assume that P and Q are relatively prime: if $f = P_1/Q_1$ and $\delta(z) = \gcd(P_1, Q_1)$, then the representation $f = P/Q$ with $P = P_1/\delta$, $Q = Q_1/\delta$ is irreducible. In addition, it is always possible to normalize Q in such a way that its constant term is 1, that is, $Q(0) = 1$. Also, Euclidean division, provides the form

$$f(z) = R(z) + \frac{\widehat{P}(z)}{Q(z)},$$

where R is a polynomial and deg $(\widehat{P}) <$ deg (Q). This transformation from P/Q to $\widehat{P}/Q$ only modifies finitely many initial values of the coefficients of f.

Observe that, with the normalization $Q(0) = 1$, we have

$$f(z) = \sum_{k=0}^{\infty} P(z)(1 - Q(z))^k,$$

where the sum is well-defined in the sense of formal power series. By simple majorization arguments ($Q(0) = 1$ implies that $1 - Q(z)$ is small for z near 0), a rational series as defined in the formal sense of Definition 8.1 always determines a function analytic at the origin.

1.1. Characterizations and elimination. A rational series in one variable can be characterized in a number of equivalent ways, by refinement of the definition, as a solution to a linear system, by the linear recurrence satisfied by its coefficient, or by the "exponential-polynomial" form of its coefficients.

THEOREM 8.1 (Rational function characterizations). *For a power series $f(z) = \sum_n f_n z^n$ in $\mathbb{C}[[z]]$, the following conditions are equivalent to rationality:* (i) Normal form: *there exist polynomials $P(z), Q(z) \in \mathbb{C}[z]$ such that*

$$f(z) = \frac{P(z)}{Q(z)},$$

with $Q(0) = 1$, and P, Q are relatively prime.

(ii) Elimination: *there exist a vector of formal power series $\mathbf{v}$, and a matrix $\mathbf{T}$ with polynomial entries and with $\det(I - \mathbf{T}(0)) \neq 0$, such that the solution $\mathbf{g}$ to the system*

$$\mathbf{g} = \mathbf{v} + \mathbf{T}\mathbf{g},$$

has $\mathbf{g}_1 = f$.

(iii) Coefficient recurrence: *there exist constants $c_1, c_2, \ldots, c_r$ such that*

$$f_{n+r} = c_1 f_{n+r-1} + c_2 f_{n+r-2} + \cdots + c_r f_n = 0,$$

for all n greater than a fixed number N_0.

(iv) Coefficients as exponential polynomials: *there exist a finite set of constants, $\{\omega_j\}_{j=1}^{c}$, and a finite set of polynomials, $\{R_j(n)\}_{j=1}^{c}$, such that*

$$f_n = \sum_{j=1}^{k} R_j(n)\omega_j^n,$$

for all n greater than a fixed number N_1.

The matrix $\mathbf{T}$ that defines a rational function via a linear system is often called a *transition matrix* or a *transfer matrix*. The form (iv) for coefficients is called an *exponential-polynomial*. Naturally, one may always assume that the ω_j are "sorted", $|\omega_1| \geq |\omega_2| \geq \cdots$, which is the key to asymptotic approximations.

Proof. (i) is equivalent to the definition of a rational power series, by the comments above.

$(i) \Longrightarrow (ii)$ results from the fact that f satisfies the 1-dimensional system $f = P + (1 - Q)f$. The converse property $(ii) \Longrightarrow (i)$ results from Cramer's solution of linear systems in terms of determinants. This provides a rational form for f with denominator $\det(I - \mathbf{T}(z))$ that is locally nonzero since, by assumption, $\det(I - \mathbf{T}(0)) \neq 0$. (Note that this condition is in particular automatically satisfied inn the frequent case where $T(0)$ is nilpotent.)

$(i) \Longrightarrow (iii)$ arises from the identity

$$0 = Q(z)f(z) - P(z),$$

upon extracting the coefficient of z^n. The converse implication $(iii) \Longrightarrow (i)$ results from translating the recurrence into an OGF equation in the standard way (multiply by z^n and sum).

$(i) \Longrightarrow (iv)$ is obtained by partial fraction expansion followed by coefficient extraction by means of standard identity

$$\frac{1}{(1 - \omega z)^r} = \sum_{n \geq 0} \binom{n + r - 1}{r - 1} \omega^n z^n,$$

where the binomial coefficient is a polynomial in n of exact degree $r - 1$. The converse implication $(iv) \Longrightarrow (i)$ results from the same identity used in the opposite direction to synthesize the function from its coefficients,

$$\sum_{n \geq 0} \binom{n + r - 1}{r - 1} \omega^n z^n = \frac{1}{(1 - \omega z)^r},$$

where the binomial coefficients $\binom{n+r-1}{r-1}$ $(r \geq 1)$ form a basis of the set of polynomials $\mathbb{C}[n]$. □

We have opted for the use of determinants as an approach to elimination in linear systems. An alternative is Gaussian elimination. The principle is well known: the algorithm takes a system of linear forms and combines them linearly, so as to eliminate all variables in succession, until a normal form, $x_i = \alpha_i$ is attained. When we discuss elimination in polynomial systems later in this chapter (Section 4), we shall encounter similarly two approaches: one, based on resultants, makes extensive use of determinants while the other, relying on Groebner bases, is somewhat reminiscent of Gaussian elimination.

1.2. Closure properties and coefficients. Rational functions satisfy several closure properties that derive rather directly from their definition or from the characterizations granted by Theorem 8.1.

THEOREM 8.2 (Rational series closure). *The set $\mathbb{C}^{\mathrm{rat}}[\![z]\!]$ of rational series is closed under the operations of sum $(f + g)$, product $(f \times g)$, quasi-inverse (defined by $f \mapsto (1 - f)^{-1}$, conditioned upon $f_0 = 0$), differentiation (∂_z), composition ($f \circ g$, conditioned upon $g_0 = 0$), and Hadamard (termwise) product,*

$$f(z) \odot g(z) = \left(\sum_n f_n z^n\right) \odot \left(\sum_n g_n z^n\right) = \sum_n (f_n \cdot g_n) z^n.$$

Proof. Only the Hadamard closure deserves a comment: it results from the characterization of coefficients of rational generating functions as exponential polynomials whose class is obviously closed under product. □

EXERCISE 1. Let $\{F_n\} = (0, 1, 1, 2, 3, 5, \ldots)$ be the Fibonacci sequence. Examine properties of the generating functions

$$F^{(m)}(z) = \sum_{n=0}^{\infty} (F_n)^m z^n.$$

EXERCISE 2. Develop a proof of closure under Hadamard products that is based on the Hadamard integral formula of p. 55

Closed form for coefficients. A combinatorial sum expression for coefficients of rational functions results from the expansion of $1/Q(z)$ as $\sum_k (1 - Q(z))^k$.

THEOREM 8.3 (Rational function coefficients). *Let $f(z)$ be a rational function with $f(z) = P(z)/Q(z)$ and Q normalized by $Q(0) = 1$. Set*

$$P(z) = \sum_{j=0}^{s} p_j z^j, \qquad 1 - Q(z) = r_1 z + r_2 z^2 + \cdots + r_m z^m.$$

and define the combinatorial sums:

$$S_n := \sum_{1k_1 + 2k_2 + \cdots + mk_m = n} \binom{k_1 + \cdots + k_m}{k_1, \ldots, k_m} (r_1^{k_1} r_2^{k_2} \cdots r_m^{k_m}).$$

The coefficients of f are expressible in finite terms from the S_n:

$$[z^n] \frac{P(z)}{Q(z)} = \sum_{j=0}^{s} p_j S_{n-j}.$$

Proof. The multinomial expansion gives

$$\begin{aligned} \frac{1}{Q(z)} &= \frac{1}{1 - (1 - Q(z))} \\ &= \sum_{k_1, k_2, \ldots, k_m} \binom{k_1 + \cdots + k_m}{k_1, \ldots, k_m} (r_1^{k_1} r_2^{k_2} \cdots r_m^{k_m}) z^{k_1 + 2k_2 + \cdots + mk_m}, \end{aligned}$$

and it suffices to multiply this expansion by the numerator $P(z)$. □

As a consequence, the coefficients of P/Q are expressible as multinomial sums of multiplicity at most $m - 1$ (in fact only the number of nonzero monomials in P matters). (Multiplicity is also called "index" in Comtet's book [**26**] that provides many interesting examples of nontrivial expansions.) Note that, by "Fatou's Lemma", if all the coefficients $[z^n]f(z)$ are integers, then the p_j and r_j are themselves integers. (See the discussion in [**87**, p. 264].)

EXAMPLE 1. *Fibonacci numbers and binomial coefficients.* The generating function of Fibonacci numbers, when expanded according to Theorem 8.3 leads to

$$F_{n+1} = [z^n] \frac{1}{1 - z - z^2} = \sum_{j \geq 0} \binom{n - j}{j},$$

so that Fibonacci numbers are sums of ascending diagonals of Pascal's triangle. Naturally, infinitely many variants exist when expanding a rational function. For instance, one has also

$$\frac{1}{1-(z+z^2)} = \frac{1}{1-z}\frac{1}{1-\frac{z^2}{1-z}} = \frac{1}{1-z^2}\frac{1}{1-\frac{z}{1-z^2}} = \frac{1}{1-z\frac{1-z^2}{1-z}},$$

leading to a trivial variety of binomial forms. More generally, compositions with largest summand $\leq m$ have an OGF that is

$$\frac{1}{1-(z+z^2+\cdots z^m)} = \frac{1}{1-z\frac{1-z^m}{1-z}} = \frac{1-z}{1-2z+z^{m+1}},$$

and corresponding expansions lead to sums that are of index $(m-1)$, 2, and 2, respectively. □

EXERCISE 3. The OGF of an ascending line in Pascal's triangle, $G_n = \sum_j \binom{n-dj}{j}$, is a rational function.

It is a common but often fruitless exercise in combinatorial analysis to show "directly" equivalence between such combinatorial sums. In fact, as the example above illustrates, elementary combinatorial identities are often nothing but the image (in a world with little transparent algebraic structure) of simple algebraic identities between generating functions (that live in a world with a strong structure).

2. Analysis of rational functions

In principle, the asymptotic analysis of coefficients of a rational function is "easy" given the exponential–polynomial form. However, in most applications of interest, rational functions are only given implicitly as solutions to linear systems. This confers a great value to criteria that ensure unicity and/or simplicity of the dominant pole. Accordingly, the bulk of this section is devoted to a brief exposition of Perron-Frobenius theory that covers adequately the case of positive linear systems.

2.1. General rational functions. The fact that coefficients of rational series are expressible as exponential polynomials yields an asymptotic equivalent. The ω_j satisfy $\omega_j = (\alpha_j)^{-1}$, where $\alpha_1, \alpha_2, \ldots$ are the poles of $f(z)$, that is to say, the zeros of the denominator of f. The formula simplifies as soon as there is a unique number in $\{\omega_j\}$, say ω_1, that dominates the other ones in absolute value.

THEOREM 8.4 (Rational function asymptotics). *Let $f = P/Q$ be a rational function where $\gcd(P,Q) = 1$. Assume that f has a unique dominant pole, that is, the zeros $\{\alpha_j\}$ of the denominator polynomial $Q(z)$ satisfy $|\alpha_1| < |\alpha_2| \leq |\alpha_3| \leq \cdots$, then ($\epsilon > 0$ being arbitrary)*

$$[z^n]f(z) = R_1(n)\alpha_1^{-n} + O((|\alpha_2| - \epsilon)^{-n},$$

where $R_1(n)$ is a polynomial.

If $f(z)$ has several dominant poles, $|\alpha_1| = |\alpha_2| = \cdots = |\alpha_r| < |\alpha_{r+1}| \leq \cdots$, then ($\epsilon > 0$ being arbitrary)

$$f_n = \sum_{j=1}^{r} R_j(n)\alpha_j^{-n} + O((|\alpha_{r+1}| - \epsilon)^{-n}),$$

where the $R_j(n)$ are polynomials. The degree of each R_j equals the order of the pole of $f(z)$ at α_i minus one.

Proof. Immediate from Theorem 8.1 □

Observe that, if $f(z)$ has nonnegative coefficients, $f(z) \in \mathbb{R}_{\geq 0}[\![z]\!]$, then α_1 must at least be real and positive by Pringsheim's theorem. Unicity of a zero of smallest modulus for $Q(z)$ is equivalent to unicity of dominant singularity for $f(z)$ and this is the simplest case for analysis. The prototypical application is provided by the Fibonacci numbers $F_0 = 0$, $F_1 = 1$, $F_{n+2} = F_{n+1} + F_n$ with OGF

$$f(z) = \frac{z}{1 - z - z^2}.$$

The poles are at $1/\phi$ and $1/\overline{\phi}$, where

$$\phi = \frac{1 + \sqrt{5}}{2}, \qquad \overline{\phi} = \frac{1 - \sqrt{5}}{2},$$

and one has $F_n \sim \phi^n/\sqrt{5}$.

EXAMPLE 2. *Compositions into finite summands.* The OGF of integer compositions having summands in the finite set $S = \{s_1, \ldots, s_m\}$ with $\gcd(\{s_j\}) = 1$ is

$$C(z) = \prod_{j=1}^{m} \frac{1}{1 - \sum_j z^{s_j}}.$$

The characteristic polynomial $q(z) = \sum z^{s_j}$ has positive coefficients, so that there exists a unique positive ρ satisfying $q(\rho) = 1$ and additionally one has $q'(\rho) > 0$. Also all other roots of q are of modulus strictly larger than ρ (by positivity of q or Pringsheim's theorem combined with the gcd assumption). There results that

$$[z^n]C(z) = [z^n]\frac{1}{q(\rho) - q(z)} \sim \frac{1}{q'(\rho)(\rho - z)} \sim \frac{1}{\rho q'(\rho)}\rho^{-n},$$

since the dominant pole at ρ is simple. □

The case where several dominant singularities are present can also be treated easily as soon as one of them has a higher multiplicity, as this singles out a larger contribution in the coefficients' asymptotics.

EXAMPLE 3. *Denumerants.* The OGF of integer partitions with summands in the set $S = \{s_1, \ldots, s_m\}$ where $\gcd(\{s_j\}) = 1$ is

$$D(z) = \prod_{j=1}^{m} \frac{1}{1 - z^{s_j}}.$$

This GF has poles at $z = 1$ and at roots of unity, but only the pole at $z = 1$ attains multiplicity m, with

$$D(z) \underset{z \to 1}{\sim} \frac{1}{\sigma} \frac{1}{(1 - z)^m}, \qquad \sigma = \prod_{j=1}^{m} s_j.$$

There results the estimate of the number of denumerants $[z^n]D(z) \sim \sigma^{-1}\binom{n+m-1}{m-1}$, that is,

$$[z^n]D(z) \sim \left(\prod_{j=1}^{m} s_j\right)^{-1} \frac{n^{m-1}}{(m-1)!},$$

which is due to Schur. □

In the case when there exist several dominant singularities possessing the same multiplicity, then fluctuations appear; refer to the discussion of Chapter 4. For a general linear recurrence over $\mathbb{Q}$, equivalently for a GF in $\mathbb{Q}(z)$, it is for instance only known that the set $\Omega_f = \{n \mid f_n = 0\}$ is a finite union of (finite or infinite) arithmetic progressions—this constitutes the Skolem-Mahler-Lech theorem. However, Ω_f is not known to be computable and it is not even known whether the property $\Omega_f = \emptyset$ is decidable. The function

$$\frac{1}{1-\frac{6}{5}z+z^2} = 1 + 1.20z + 0.44z^2 - 0.67z^3 - 1.24z^4 - 0.82z^5 + 0.25z^6 + \cdots,$$

that is built on the Pythagorean triple $\langle 3,4,5\rangle$ illustrates some of the hardships. We have

$$f_n = \frac{\sin(n+1)\varphi}{\sin\varphi}, \qquad \varphi = \arccos\frac{3}{5},$$

and, for instance, the way f_n approaches its extremes $\pm\frac{5}{4} = \pm\frac{1}{\sin\varphi}$ depends on how well multiples of φ approximate multiples of π, that is, on deep arithmetic properties of φ/π. Such pathological situations, though possibly present amongst general rational functions, hardly ever occur in analytic combinatorics.

EXERCISE 4. Examine empirically the sign pattern of coefficients of the rational functions

$$g_r(z) = \frac{1}{1-z} - \left(z\frac{d}{dz}\right)^r\left(\frac{5}{4(1-z)} - f(z)\right), \text{ i.e., } g_{r,n} = 1 - n^r(\frac{5}{4} - f_n).$$

2.2. Positive rational functions and Perron-Frobenius theory. For rational functions, positivity coupled with some ancillary conditions entails a host of important properties, like unicity of the dominant singularity. Such facts result from the classical Perron-Frobenius theory of nonnegative matrices that we now summarize.

Consider first rational functions of the special form

$$f(z) = \frac{1}{1-S(z)},$$

for $S(z)$ a polynomial. It is assumed that S has no constant term ($S(0)=0$), so that $f(z)$ is properly defined as a formal power series. Assume next nonnegativity of the coefficients of S; this entails existence of a dominant real pole that is simple. If additionally $S(z)$ is not a polynomial in z^d for some $d \geq 2$, then, as is easy to see, there is uniqueness (and simplicity) of dominant singularity. As we show now, similar properties hold for systems: the conditions are those of properness and positivity coupled with a new combinatorial notion of "irreducibility". Granted these conditions, the analysis suitably extends what we just saw of the scalar case.

2.2.1. *Perron-Frobenius theory of nonnegative matrices.* The properties of positive and of nonnegative matrices have been superbly elicited by Perron [**77**] in 1907 and by Frobenius [**43**] in 1908–1912. The corresponding theory has far-reaching implications: it lies at the basis of the theory of finite Markov chains and it extends to positive operators in infinite-dimensional spaces [**60**].

For a square matrix $M \in \mathbb{R}^{m\times m}$, the *spectrum* is the set of its *eigenvalues*, that is, the set of λ such that $\lambda I - M$ is not invertible (i.e., not of full rank), where I is the unit matrix with the appropriate dimension. A *dominant eigenvalue* is one of largest modulus.

For M a scalar matrix of dimension $m \times m$ with nonnegative entries, a crucial rôle is played by the *dependency graph*; this is the (directed) graph that has vertex set $V = \{1\,..\,m\}$ and edge set containing the directed edge $(a \to b)$ iff $M_{a,b} \neq 0$. The

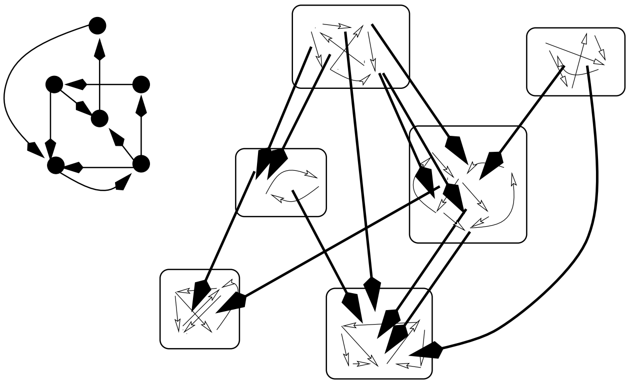


FIGURE 2. The irreducibility conditions of Perron-Frobenius theory. Left: a strongly connected digraph. Right: a weakly connected digraph that is not strongly connected is a collection of strongly connected components related by a directed acyclic graph.

reason for this terminology is the following: Let M represent the linear transformation $\left\{y_i^\star = \sum_j M_{i,j} y_j\right\}_i$; then, a nonzero entry $M_{i,j}$ means that $y_i^\star$ depends effectively on y_j, a fact translated by the directed edge $(i \to j)$.

Two notions are essential, irreducibility and aperiodicity (the terms are borrowed from Markov chain theory and matrix theory).

Irreducibility. The matrix M is called *irreducible* if its dependency graph is strongly connected (i.e., any two points are connected by a path). By considering only simple paths, it is then seen that irreducibility is equivalent to the condition that $(I + M)^m$ has all its entries that are strictly positive. See Figure 2 for a graphical rendering of irreducibility and for the general structure of a (weakly connected) digraph.

Periodicity. A strongly connected digraph G is *periodic* with parameter d iff all its cycles have a length that is a multiple of d. In that case, the graph decomposes into cyclically arranged layers: the vertex set V can be partitioned into d classes, $V = V_0 \cup \cdots \cup V_{d-1}$, in such a way that the edge set E satisfies

$$E \subseteq \bigcup_{i=0}^{d-1} (V_i \times V_{(i+1) \bmod d}). \tag{4}$$

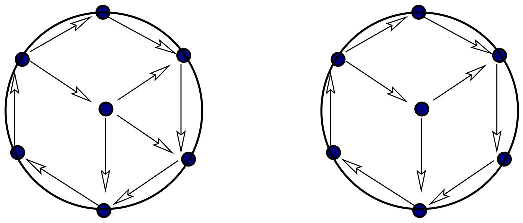
The maximal possible d is called the *period*. (For instance a directed 10-cycle is periodic with parameter $d = 1, 2, 5, 10$ and the period is 10.) If no decomposition exists with $d \geq 2$, so that the period has the trivial value 1, then the graph and all the matrices that admit it as their dependency graph are called *aperiodic*. See Figure 3.

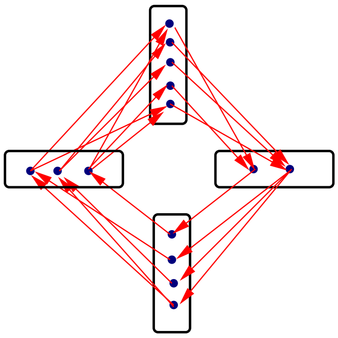
THEOREM 8.5 (Perron-Frobenius theory). *Let M be a matrix that is assumed to be irreducible, i.e., its dependency graph is strongly connected.*

(i) *If M has (strictly)* positive elements, *then its eigenvalues can be ordered in such a way that*

$$\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots ,$$

and M has a unique dominant eigenvalue; this eigenvalue is positive and simple.





$$M = \begin{bmatrix} 0&0&0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&0&0&1&1&0\\ 1&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0&0&0&0&0&0\\ 1&0&0&1&1&0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

FIGURE 3. The aperiodicity conditions of Perron-Frobenius theory. Top: an aperiodic digraph (left) and a periodic digraph (right). Bottom: A digraph of period $d = 4$ (left) corresponding to a matrix M (right). For irreducible matrices, the (combinatorial) property of the graph to have period d is equivalent to the (analytic) property of the existence of d dominant eigenvalues and it implies a rotational invariance of the spectrum. Here, the characteristic polynomial is $z^6(z^8 - 4z^4 + 2)$. (The spectrum consists of $(2 \pm 2^{1/2})^{1/4}$ and their four conjugates, as well as 0.)

(ii) *If M has* nonnegative elements, *then its eigenvalues can be ordered in such a way that*

$$\lambda_1 = |\lambda_2| = \cdots = |\lambda_d| > |\lambda_{d+1}| \geq |\lambda_{d+2}| \geq \cdots,$$

and each of the dominant eigenvalues is simple with λ_1 positive.

Furthermore, the quantity d is precisely equal to the period of the dependency graph. If $d = 1$, in particular, then there is unicity of the dominant eigenvalue. If $d \geq 2$, the whole spectrum is invariant under the set of transformations

$$\lambda \mapsto \lambda e^{2ij\pi/d}, \qquad j = 0, 1, \ldots, d-1.$$

Periodicity also means that the existence of paths of length n between any given pair of nodes $\langle i, j\rangle$ is constrained by the congruence class $n \bmod d$. A contrario, aperiodicity entails the existence, for all n sufficiently large, of paths of length n connecting $\langle i, j\rangle$. From the definition, a matrix M with period d has, up to simultaneous permutation of its rows and columns, a cyclic block structure

$$\begin{pmatrix} 0 & \boxed{M_{0,1}} & 0 & \cdots & 0 \\ 0 & 0 & \boxed{M_{1,2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \boxed{M_{d-2,d-1}} \\ \boxed{M_{d-1,0}} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where the blocks $M_{i,i+1}$ are reflexes of the connectivity between V_i and V_{i+1} in (4).

For short, one says that a matrix is positive (resp. nonnegative) if all its elements are positive (resp. nonnegative). Here are two useful turnkey results, Corollaries 8.1 and 8.2.

COROLLARY 8.1. *Any one of the following conditions suffices to guarantee the existence of a unique dominant eigenvalue of a nonnegative matrix T:*

(i) *T has (strictly) positive entries;*
(ii) *T is such that, some power T^s is (strictly) positive;*
(iii) *T is irreducible and at least one diagonal element of T is nonzero;*
(iv) *T is irreducible and the dependency graph of T is such that there exist at least two paths from the same source to the same destination that are of relatively prime lengths.*

Proof. The proof makes an implicit use of the correspondence between terms in coefficients of matrix products and paths in graphs (see below, Section 3.3 for more).

Sufficiency of condition (i) results directly from Case (i) of Theorem 8.5.

Condition (ii) immediately implies irreducibility. Unicity of the dominant eigenvalue (hence aperiodicity) results from Perron-Frobenius properties of M^s, by which $\lambda_1^s > |\lambda_2|^s$. (Also, by elementary graph combinatorics, one can always take the exponent s to be at most the dimension m.)

By basic combinatorics of paths in graphs, Conditions (iii) and (iv) imply Condition (ii). □

2.2.2. *Positive rational functions.* The importance of Perron-Frobenius theory and of its immediate consequence, Corollary 8.1, stems from the fact that uniqueness of the dominant eigenvalue is usually related to a host of analytic properties of generating functions as well as probabilistic properties of structures. In particular, as we shall see in the next section, several combinatorial problems (like automata or paths in graphs) can be reduced to the following case.

COROLLARY 8.2. *Consider the matrix*

$$F(z) = (I - zT)^{-1},$$

where T, called the "transition matrix", is a scalar nonnegative matrix. It is assumed that T is irreducible. Then each entry $F_{i,j}(z)$ of $M(z)$ has a radius of convergence ρ that coincides with the smallest positive root of the determinantal equation

$$\Delta(z) := \det(I - zT) = 0.$$

Furthermore, the point ρ is a simple pole of any $F_{i,j}(z)$.

In addition, if T is aperiodic or if it satisfies any of the conditions of Corollary 8.1, then all singularities other than ρ are strictly dominated in modulus by ρ.

Proof. Define first (as in the statement) $\rho = 1/\lambda_1$, where λ_1 is the eigenvalue of T of largest modulus that is guaranteed to be simple by assumption of irreducibility and by Perron-Frobenius properties. Next, the relations induced by $F = I + zTF$, namely,

$$F_{i,j}(z) = \delta_{i,j} + z \sum_k T_{i,k} F_{k,j}(z),$$

together with positivity and irreducibility entail that the $F_{i,j}(z)$ must all have the same radius of convergence r. Indeed, each F_{ij} depends positively on all the other ones (by irreducibility) so that any infinite value of an entry in the system must propagate to all the other ones.

The characteristic polynomial

$$\Delta(z) = \det(I - zT),$$

has roots that are inverses of the eigenvalues of T and $\rho = 1/\lambda_1$ is smallest in modulus. Thus, since Δ is the common denominator to all the $F_{i,j}(z)$, poles of any $F_{i,j}(z)$ can only be included in the set of zeros of this determinant, so that the inequality $r \geq \rho$ holds.

It remains to exclude the possibility $r > \rho$, which means that no “cancellations” with the numerator can occur at $z = \rho$. The argument relies on finding a positive combination of some of the $F_{i,j}$ that *must* be singular at ρ. We offer two proofs, each of interest in its own right: one (a) is conveniently based on the Jacobi trace formula, the other (b) is based on supplementary Perron–Frobenius properties.

(a) Jacobi’s trace formula for matrices [**48**, p. 11],

$$\det \circ \exp = \exp \circ \operatorname{Tr} \qquad \text{or} \qquad \log \circ \det = \operatorname{Tr} \circ \log \tag{5}$$

generalizes the scalar identities[1] $e^a e^b = e^{a+b}$ and $\log ab = \log a + \log b$. Here we have (for z small enough)

$$\begin{aligned} \operatorname{Tr} \log(I - zT)^{-1} &= \sum_i \sum_{n \geq 1} M_{i,i,n} \frac{z^n}{n} \\ &= \log(\det(I - zT)^{-1}, \end{aligned}$$

where the first line results from expansion of the logarithm and the second line is an instance of the trace formula. Thus, by differentiation, the sum $\sum_i M_{i,i}(z)$ is seen to be singular at $\rho = 1/\lambda_1$ and we have established that $r = \rho$.

(b) Alternatively, let v_1 be the eigenvector of T corresponding to λ_1. Perron-Frobenius theory also teaches us that, under the irreducibility and aperiodicity conditions, the vector v_1 has all its coordinates that are nonzero. Then the quantity

$$(1 - zT)^{-1} v_1 = \frac{1}{1 - z\lambda_1} v_1$$

is certainly singular at $1/\lambda_1$. But it is also a linear combination of the $F_{i,j}$’s. Thus at least one of the entries of F (hence all of them by the discussion above) must be singular at $\rho = 1/\lambda_1$. Therefore, we have again $r = \rho$.

Finally, under the additional assumption that T is aperiodic, there follows from Perron-Frobenius theory that $\rho = 1/\lambda_1$ is well-separated in modulus from all other singularities F. □

[1] The Jacobi trace formula is readily verified when the matrix is diagonalizable, and from there, it can be extended to all matrices by an algebraic “density” argument.

It is interesting to note that several of these arguments will be recycled when we discuss the harder problem of analysing coefficients of positive algebraic functions in Section 5.2.

EXERCISE 5. The de Bruijn matrix $T \in \mathbb{R}^{2^\ell \times 2^\ell}$ is essential in problems related to occurrences of patterns in random strings [**39**]. Its entries are given by

$$T_{i,j} = \frac{1}{2}[\![(j = 2i \bmod 2^\ell) \text{ or } (j = 2i+1 \bmod 2^\ell)]\!].$$

Prove that it has a unique dominant eigenvalue. [Hint: consider a suitable graph with vertices labelled by binary strings of length ℓ.]

We next proceed to show that properties of the Perron-Frobenius type even extend to a large class of linear systems of equations that have nonnegative polynomial coefficients. Such a case is important because of its applicability to transfer matrices; see Section 3.3 below.

Some definitions extending the ones of scalar matrices must first be set. A polynomial

$$p(z) = \sum_j c_j z^{e_j}, \qquad \text{every } c_j \neq 0,$$

is said to be primitive if the quantity $\delta = \gcd(\{e_j\})$ is equal to 1; it is imprimitive otherwise. Equivalently, $p(z)$ is imprimitive iff $p(z) = q(z^\delta)$ for some *bona fide* polynomial q and some $\delta > 1$. Thus, $z, 1+z, z^2+z^3, z+z^4+2z^8$ are primitive while $1, 1+z^2, z^3+z^6, 1+2z^8+5z^{12}$ are not.

DEFINITION 8.2. *A linear system with polynomial entries,*

$$f(z) = v(z) + T(z)f(z) \tag{6}$$

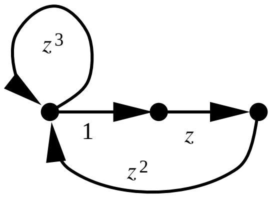
where $T \in \mathbb{R}[z]^{r\times r}$, $v \in \mathbb{R}[z]^r$, and $f \in \mathbb{R}[z]^r$ the vector of unknowns is said to be:

- (a) rationally proper (r–proper) *if $T(0)$ is nilpotent, meaning that $T(0)^r$ is the null matrix;*
- (b) rationally nonnegative (r–nonnegative) *if each component $v_j(z)$ and each matrix entry $T_{i,j}(z)$ lies in $\mathbb{R}_{\geq 0}[z]$;*
- (c) rationally irreducible (r–irreducible) *if $(I + T(z))^r$ has all its entries that are nonzero polynomials.*
- (d) rationally aperiodic (r–periodic) *if at least one diagonal entry of some power $T(z)^e$ is a primitive polynomial.*

It is again possible to visualize these properties of matrices by drawing a directed graph whose vertices are labelled $1, 2, \ldots, r$, with the edge connecting i to j that is weighted by the entry $T_{i,j}(z)$ of matrix $T(z)$. Properness means that all sufficiently long paths (and all cycles) must involve some positive power of z— it is a condition satisfied in well-founded combinatorial problems; irreducibility means that the dependency graph is strongly connected by paths involving edges with nonzero polynomials. Periodicity means that all closed paths involve weights that are polynomials in some z^e for some $e > 1$.

For instance, if W is a matrix with positive entries, then zW is r–irreducible and r–aperiodic, while z^3W is r–periodic. The matrix $T = \begin{pmatrix} z & z^3 \\ 1 & 0 \end{pmatrix}$ is r–proper, r–irreducible, and r–aperiodic, since $T^2 = \begin{pmatrix} z^2+z^3 & z^4 \\ z & z^3 \end{pmatrix}$. The matrix $T =$

$\begin{pmatrix} z^3 & 1 & 0 \\ 0 & 0 & z \\ z^2 & 0 & 0 \end{pmatrix}$ is r–proper, but it fails to be r–aperiodic since, for instance, all cycles only involve powers of z^3, as is visible on the associated graph:



By abuse of language, we say that $f(z)$ is a solution of a linear system if it coincides with the first component of a solution vector, $f \equiv f_1$. The following theorem generalizes Corollary 8.2.

THEOREM 8.6 (Positive rational systems). (i) *Assume that a rational function $f(z)$ is a solution of a system (6) that is r–positive, r–proper, r–irreducible, and r–aperiodic. Then, $f(z)$ has a unique dominant singularity ρ that is positive, and is a simple pole; ρ is the smallest positive solution of*

$$\det(I - T(z)) = 0. \tag{7}$$

(ii) *Assume that $f(z)$ is a solution of a system that is r–positive, r–proper, and r–irreducible (but not necessarily r–aperiodic). Then, the set of dominant singularities of $f(z)$ is of the form $\{\rho_j\}_{j=0}^{d-1}$, where $\rho_0 \in \mathbb{R}_{\geq 0}$, $\rho_j/\rho_0 = \eta$ is a root of unity, and $\rho_j\eta^\ell$ is a dominant singularity for all $\ell = 0, 1, 2, \ldots$. In addition, each ρ_j is a simple pole.*

Proof. Consider first Case (i). For any fixed $x > 0$, the matrix $T(x)$ satisfies the Perron Frobenius conditions, so that it has a maximal positive eigenvalue $\lambda_1(x)$ that is simple. More information derives from the introduction of matrix norms[2]. The spectral radius of an arbitrary matrix A is defined as

$$\sigma(A) = \max_j\{|\lambda_j|\}, \tag{8}$$

where the set $\{\lambda_j\}$ is the set of eigenvalues of A (also called spectrum). Spectral radius and matrix norms are intimately related since

$$\sigma(A) = \lim_{n\to+\infty} (||A^n||)^{1/n}.$$

In particular, this relation entails that the spectral radius is an increasing function of matrix entries: for nonnegative matrices, if $A \leq B$ in the sense that $A_{i,j} \leq B_{i,j}$ (for all i, j), then $\sigma(A) \leq \sigma(B)$; if $A < B$ in the sense that $A_{i,j} < B_{i,j}$ (for all i, j), then $\sigma(A) < \sigma(B)$. (To see the last inequality, note the existence of $\epsilon > 0$ such that $A \leq (1-\epsilon)B$.)

Returning to the case at hand, equation (8) and the surrounding remarks imply that the spectral radius $\sigma(T(x))$, which also equals $\lambda_1(x)$ for positive x, satisfies

$$\lambda_1(0) = 0, \qquad \lambda_1(x) \text{ strictly increasing}, \qquad \lambda_1(+\infty) = +\infty.$$

(The first condition reflects properness, the second one is a consequence of irreducibility, and the last one derives from simple majorizations.) In particular, the equation $\lambda_1(x) = 1$ admits a unique root ρ on $(0, +\infty)$. (Notice that $\lambda_1(x)$ is a real branch of the algebraic

[2]A matrix norm $||.||$ satisfies: $||A|| = 0$ implies $A = 0$; $||cA|| = |c| \cdot ||A||$; $||A + B|| \leq ||A| + ||B||$; $||A \times B|| \leq ||A|| \cdot ||B||$.

curve $\det(\lambda I - T(x)) = 0$ that dominates all other branches in absolute value for $x > 0$. There results from the general theory of algebraic functions, see Section 5, that $\lambda_1(x)$ is analytic at every point $x > 0$.)

There remains to prove that: (a) ρ is at most a simple pole of $f(z)$; (b) ρ is actually a pole; (c) there are no other singularities of modulus equal to ρ.

Fact (a) amounts to the property that ρ is a simple root of the equation $\lambda(\rho) = 1$, that is, $\lambda'(\rho) \neq 0$. (To prove $\lambda'(\rho) \neq 0$, we can argue a contrario. First derivatives $\lambda'(\rho), \lambda''(\rho)$, etc, cannot be zero till some *odd* order inclusively since this would contradict the increasing character of $\lambda(x)$ around ρ along the real line. Next, if derivatives till some *even* order ≥ 2 inclusively were zero, then we would have by the local analytic geometry of $\lambda(z)$ near ρ some complex value z_1 satisfying: $|\lambda(z_1)| = 1$ and $|z_1| < \rho$; but for such a value z_1, by irreducibility and aperiodicity, for some exponent e, the entries of $T(z_1)^e$ would be all strictly dominated in absolute value by those of $T(\rho)^e$, hence a contradiction.) Then, $\lambda'(\rho) \neq 0$ holds and by virtue of

$$\det(I - T(z)) = (1 - \lambda_1(z)) \prod_{j \neq 1} (1 - \lambda_j(z)) = (1 - \lambda_1(z)) \frac{\det(I - T(z))}{1 - \lambda_1(z)},$$

the quantity ρ is only a simple root of $\det(I - T(z))$.

Fact (b) means that no "cancellation" may occur at $z = \rho$ between the numerator and the denominator given by Cramer's rule. It derives from an argument similar to the one employed for Corollary 8.2. Fact (c) derives from aperiodicity and the Perron-Frobenius properties. $\square$

3. Combinatorial applications of rational functions

Rational functions occur as generating functions of well-recognized classes of enumerative problems. We examine below the case of regular specifications and regular languages (Subsections 3.1 and 3.2) that are closely related to transfer matrix methods and finite state models or automata (Subsection 3.3). Local constraints in permutations represent a direct application of transfer matrix methods (Subsection 3.4). Lattice paths lead to an extension of the regular framework to infinite alphabets and infinite grammars, revealing interesting connections with the theory of continued fractions (Subsection 3.5). For instance, the classic continued fractions identity (originally due to Gauß),

$$\sum_{s=0}^{\infty} (1 \cdot 3 \cdots (2n-1))\, z^{2n} \frac{1}{1 - \cfrac{1 \cdot z^2}{1 - \cfrac{2 \cdot z^2}{\ddots}}},$$

expresses combinatorially a very "regular" decomposition of involutive permutations and has implications on the physics of random interconnection networks.

3.1. Regular specifications. A combinatorial specification is said to be *regular* if it is nonrecursive ("iterative", see Chapter 1) and it involves only the constructions of Atom, Union, Product, and Sequence. We consider here unlabelled structures. Since the operators associated to these constructions are all rational, it follows that the corresponding OGF is rational. The OGF can be then systematically obtained from the specification by the symbolic methods of Chapter 1.

For instance, the OGF of the class $\mathcal{G}_h$ of general Catalan trees of height at most h is defined by the recurrence

$$\mathcal{G}_0 = Z;\ \mathcal{G}_{h+1} = Z \times \mathfrak{S}\{\mathcal{G}_h\}.$$

Accordingly, the OGF's are

$$G_0(z) = z,\ G_1(z) = \frac{z}{1-z},\ G_2(z) = \cfrac{z}{1 - \cfrac{z}{1-z}}, \ldots,$$

and $G_h(z)$ is none other than the hth convergent in the continued fraction expansion of the Catalan GF:

$$\frac{1}{2}\left(1 - \sqrt{1-4z}\right) = \cfrac{z}{1 - \cfrac{z}{1 - \cfrac{z}{1 - \cfrac{z}{\ddots}}}}.$$

The interesting connections with Chebyshev polynomials and Mellin transform techniques —the expected height of a tree of size n turns out to be asymptotic to $\sqrt{\pi n}$— are detailed in Chapter 7.

In a similar vein, the class $\mathcal{C}^{[h]}$ of integer compositions whose summands are at most h is

$$\mathcal{C}^{[h]} = \mathfrak{S}\{Z \times \mathfrak{S}\{Z, \text{card} < h\}\} \qquad \text{so that} \qquad C^{[h]}(z) = \frac{1}{1 - z - z^2 - \cdots - z^h}.$$

The interesting asymptotics is again discussed in Chapter 7 in connection with Mellin transform asymptotics. The largest summand in a random composition of n turns out to have expectation about $\log_2 n$; see [**49**] for a general discussion.

EXERCISE 6. The OGF of integer compositions and of integer partitions with summands constrained to be either in number at most m or of size at most s is rational.

3.2. Regular languages. The name "regular specification" has been chosen in order to be in agreement with the notions of regular expression and regular language from formal language theory. These two concepts are now defined formally.

A *language* is a set of words over some fixed alphabet $\mathcal{A}$. The structurally simplest (yet nontrivial) languages are the *regular languages* that can be defined in a variety of ways: by regular expressions and by finite automata, either deterministic or nondeterministic.

DEFINITION 8.3. *The category* RegExp *of* regular expressions *is defined as the category of expressions that contains all the letters of the alphabet* $(a \in \mathcal{A})$ *as well as the empty symbol* ϵ*, and is such that, if* $R_1, R_2 \in \mathsf{RegExp}$*, then the formal expressions* $R_1 \cup R_2$*,* $R_1 \cdot R_2$ *and* $R_1^\star$ *are regular expressions.*

Regular expressions are meant to specify *languages*. The language $\mathcal{L}(R)$ denoted by a regular expression R is defined inductively by the rules: (i) $\mathcal{L}(R) = \{a\}$ if R is the letter $a \in \mathcal{A}$ and $\mathcal{L}(R) = \{\epsilon\}$ (with ϵ the empty word) if R is the symbol ϵ; (ii) $\mathcal{L}(R_1 \cup R_2) = \mathcal{L}(R_1) \cup \mathcal{L}(R_2)$ (with $\cup$ the set-theoretic union); (iii) $\mathcal{L}(R_1 \cdot R_2) = \mathcal{L}(R_1) \cdot \mathcal{L}(R_2)$ (with $\cdot$ the concatenation of words extended to sets); (iv) $\mathcal{L}(R_1^\star) = \{\epsilon\} + \mathcal{L}(R_1) + \mathcal{L}(R_1) \cdot \mathcal{L}(R_1) + \cdots$. A language is said to be a *regular language* if it is specified by a regular expression.

A language is a *set* of words, but a word $w \in \mathcal{L}(R)$ may be parsable in several ways according to R. More precisely, one defines the *ambiguity coefficient* (or *multiplicity*) of w

with respect to the regular expression R as the number of parsings, written $\kappa(w) = \kappa_R(w)$. In symbols, we have

$$\kappa_{R_1 \cup R_2}(w) = \kappa_{R_1}(w) + \kappa_{R_2}(w), \quad \kappa_{R_1 \cdot R_2}(w) = \sum_{u \cdot v = w} \kappa_{R_1}(u)\kappa_{R_2}(v),$$

with natural initial conditions ($\kappa_a(b) = \delta_{a,b}$, $\kappa_\epsilon(w) = \delta_{\epsilon,w}$), and with the definition of $\kappa_{R^\star(w)}$ taken as induced by the definition of $R^\star$ via unions and products, namely,

$$\kappa_{R^\star}(w) = \delta_{\epsilon,w} + \sum_{j=1}^{\infty} \kappa_{R^j}(w).$$

As such, $\kappa(w)$ lies in the completed set $\mathbb{N} \cup \{+\infty\}$. We shall only consider here regular expressions R that are *proper,* in the sense that $\kappa_R(w) < +\infty$. It can be checked that this condition is equivalent to requiring that no $S^\star$ with $\epsilon \in \mathcal{L}(S)$ enters in the inductive definition of the regular expression R. (This condition is substantially equivalent to the notion of well-founded specification in Chapter 1.) A regular expression R is said to be *unambiguous* iff for all w, we have $\kappa_R(w) \in \{0, 1\}$; it is said to be *ambiguous,* otherwise.

Given a language $L = \mathcal{L}(R)$, we are interested in two enumerating sequences

$$L_{R,n} = \sum_{|w|=n} \kappa_R(w), \qquad L_n = \sum_{|w|=n} \mathbf{1}_{w \in L},$$

corresponding to the counting of words in the language, respectively, with and without multiplicities. The corresponding OGF's will be denoted by $L_R(z)$ and $L(z)$. (Note that, for a given language, the definition of L is intrinsic, while that of L_R is dependent on the particular expression R that describes the language.) We have the following.

PROPOSITION 8.1 (Regular expression counting). *Given a regular expression R assumed to be of finite ambiguity, the ordinary generating function $L_R(z)$ of the language $\mathcal{L}(R)$, counting* with multiplicity*, is given by the inductive rules:*

$$\epsilon \mapsto 1, \quad a \mapsto z, \quad \cup \mapsto +, \quad \cdot \mapsto \times, \quad \star \mapsto (1 - (.))^{-1}.$$

In particular, if R is unambiguous, then the ordinary generating function satisfies $L_R(z) = L(z)$ and is given directly by the rules above.

Proof. Formal rules associate to any proper regular expression R a specification $\mathcal{R}$:

$$\begin{aligned} &\epsilon \mapsto 1 \text{ (the empty object)}, \quad a \mapsto Z_a \ (Z_a \text{ an atom}), \\ &\cup \mapsto +, \cdot \mapsto \times, \quad \star \mapsto \mathfrak{S}\{.\} \end{aligned}$$

It is easily recognized that this mapping is such that $\mathcal{R}$ generates exactly the collection of all parsings of words according to R. The translation rules of Chapter 1 then yield the first part of the statement. The second part follows since $L(z) = L_R(z)$ whenever R is unambiguous. □

The technique implied by Proposition 8.1 has already been employed silently in earlier chapters. For instance, the regular expression

$$W^{(3)} = (\epsilon + a + aa) \cdot (b \cdot (\epsilon + a + aa))^\star$$

generates unambiguously all binary words over $\{a, b\}$ without 3-runs of the letter a; see Chapter 1. The generating function is then

$$W^{(3)}(z) = (1 + z + z^2)\,\frac{1}{1 - z(1 + z + z^2)} = \frac{1 + z + z^2}{1 - z - z^2 - z^3} = \frac{1 - z^3}{1 - 2z + z^4}.$$

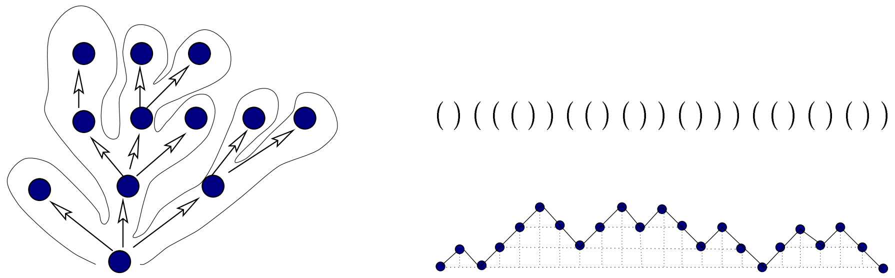


FIGURE 4. A tree of height 3, the corresponding parenthesis encoding of nesting depth 3, and the lattice path representation of height 3.

More generally, the OGF of binary words without k runs (of the letter a) is found by similar devices to be

$$W^{(k)}(z) = \frac{1 - z^k}{1 - 2z + z^{k+1}}.$$

Consider next parenthesis systems where a plays the role of the open parenthesis '(' and b is the closing parenthesis ')'. The collection of words with parenthesis nesting at most 3 is described unambiguously by

$$P^{(3)} = \left(a\,(a(ab)^\star b)^\star\, b\right)^\star,$$

so that the OGF is

$$P^{(3)}(z) = \cfrac{1}{1 - \cfrac{z^2}{1 - \cfrac{z^2}{1 - z^2}}} = \frac{1 - 2z^2}{1 - 3z^2 + z^4}.$$

The general version with parenthesis nesting of depth at most k corresponds bijectively to lattice paths in $\mathbb{N}\times\mathbb{N}$ of height $\le k$, and to general trees of height $\le k$. Indeed, the classical correspondence builds a lattice path that is the trace (up or down) of edge traversals; see Figure 4.

EXAMPLE 4. *Pattern occurrences in strings.* This is a simple example where counting with ambiguity is used. Fix a binary alphabet $\mathcal{A} = \{a, b\}$ and consider the random variable $\Omega_{u,n}$ that represents the number of occurrences of a fixed pattern $u \in \mathcal{A}^\star$ in a random string of $\mathcal{A}^n$. (See Chapter 7 of [**84**] for related material.) The regular expression (with $u = u_1 \cdots u_m$)

$$O_u = \mathcal{A}^\star(u_1 \cdot u_2 \cdots u_m)\mathcal{A}^\star$$

describes ambiguously the language of all the words that contain the pattern u. It is easy to see that a word $w \in \mathcal{A}^\star$ has an ambiguity coefficient (with respect to O_u) that is exactly equal to the number of occurrences of u in w. Clearly, the OGF associated to O_u (with multiplicity counted) is

$$O_u(z) = \frac{1}{1-2z} z^m \frac{1}{1-2z} = \frac{z^m}{(1-2z)^2}.$$

This gives immediately the mean number of occurrences of u assuming a uniform distribution over $\mathcal{A}^n$, namely

$$E(\Omega_{u,n}) = \frac{1}{2^n}\,[z^n]\frac{z^m}{(1-2z)^2} = (n - m + 1)2^{-m}.$$

The mean is asymptotic to $n2^{-m}$. In other words, there is on average a fraction about 2^{-m} of positions at which a pattern of length m is to be found in a random text of length n. A ultimate consequence is that naïve string-matching has expected complexity that is $\leq 2n$ on random texts of size n; see [**84**]. □

> EXERCISE 7. Analyse the moment of order 2 of the number of occurrences by relating it to the "correlations" between the pattern and its shifted versions. [Hint: relate the problem to the Guibas-Odlyzko correlation polynomial as described in [**84**].]

EXAMPLE 5. *Order statistics.* Consider an ordered alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$, where it is assumed that $a_1 < a_2 < \cdots < a_m$. Given a word $w = w_1 \cdots w_n$, the jth letter w_j is a record in w (respectively a weak record) if it is strictly larger (resp. not smaller) than all the previous letters $w_1, \ldots, w_{j-1}$. The study of records is a classical topic in statistical theory [**28**], and we are examining records in permutations of a multiset since repeated letters are allowed to compose w.

Regular expressions are well-suited to the problem. Consider first (strong) records. The collection of words such that the values of their records are $a_{j_1} < \cdots < a_{j_r}$ is described by the regular expression

$$R_{j_1,\ldots,j_r} = a_{j_1}\mathcal{A}^{\star}_{\leq j_1} a_{j_2}\mathcal{A}^{\star}_{\leq j_2} \cdots a_{j_r}\mathcal{A}^{\star}_{\leq j_r} \quad \text{where} \quad \mathcal{A}_{\leq j} = \{a_1, \ldots, a_j\}.$$

On the other hand, the product $\prod_j (\epsilon + a_j)$ generates all the words formed by an increasing sequence of distinct letters. These two observations combine: the regular expression

$$R = \bigcup_{j_1 < \cdots < j_r} R_{j_1,\ldots,j_r} = \prod_{j=1}^{m} \left(\epsilon \cup a_j \mathcal{A}^{\star}_{\leq j}\right)$$

generates all the words in $\mathcal{A}^{\star}$, where the number of records appears as the number (r) of factors different from ϵ in the full expansion of R.

Consequently, by the principles of Chapter 3, the multivariate OGF of words $R(u; x)$, with u marking the number of records, x an abbreviation for $x_1, \ldots, x_m$, and x_j the variable that marks the number of occurrences of letter a_j, is given by

$$R(u; x) = \left(1 + \frac{ux_1}{1 - x_1}\right)\left(1 + \frac{ux_2}{1 - x_1 - x_2}\right) \cdots \left(1 + \frac{ux_m}{1 - x_1 - \cdots - x_m}\right).$$

One checks that $R(1; x_1, \ldots, x_m) = (1 - x_1 - \cdots - x_m)^{-1}$ as should be. The generating function R is a huge multivariate extension of the Stirling cycle polynomials and, for instance, one has

$$[u^r]\,[x_1 x_2 \cdots x_m]\, R(u, x) = \left[\frac{m}{r}\right].$$

Assume that each letter a_j of a word in $\mathcal{A}^n$ has probability p_j and letters occur independently in words. This model is treated by the substitution $x_j \mapsto p_j z$. The mean number of records is then found as the coefficient of z^n in $\partial/\partial u R(u; x)$ taken at $x_j = zp_j, u = 1$. The asymptotic estimate results from straight singularity analysis of the pole at $z = 1$: *The mean number of records in a random word of length n with letter j chosen independently with probability p_j is asymptotic to*

$$\frac{p_1}{1} + \frac{p_2}{1 - p_1} + \cdots + \frac{p_m}{1 - p_1 - \cdots - p_{m-1}}.$$

The analysis as z tends to 1, keeping u as a parameter, shows also that the limit of $[z^n]R(u,z)$ exists. By the continuity theorem for probability generating functions described in Chapter 9, this implies: *The distribution of the number of records when n tends to ∞ converges to the (finite) law with probability generating function*

$$\left(1+\frac{up_1}{1-p_1}\right)\left(1+\frac{up_2}{1-p_1-p_2}\right)\cdots\left(1+\frac{up_{m-1}}{1-p_1-\cdots-p_{m-1}}\right)up_m.$$

Similarly, the multivariate GF

$$\widehat{R}(u;x)=\left(1+\frac{ux_1}{1-ux_1}\right)\left(1+\frac{ux_2}{1-x_1-ux_2}\right)\cdots\left(1+\frac{ux_m}{1-x_1-\cdots-ux_m}\right).$$

counts words by their number of weak records. This time, there is a double pole at $z=1$: *The number of (weak) records has a mean asymptotic to $c\cdot n$ (c a computable constant) and a limit Gaussian law.* ☐

Permutations and combinations of multisets are a classical topic in combinatorial analysis; see for instance MacMahon's book [**68**]. The corresponding statistics are of interest to computer scientists since they relate to searching in the context of sets and multisets obeying nonuniform data distributions. This last topic is for instance considered by Knuth [**59**, 1.2.10.18] and developed by Burge [**17**] in a pre-symbolic context. Such techniques have been successfully employed to analyse data structures like the ternary search tries of Bentley and Sedgewick [**9**]; see [**24**]. Prodinger has also developed a collection of studies concerning words whose letter probabilities are geometrically distributed, $p_j=(1-q)q^j$: see for instance [**55**]. In the latter case, one may legitimately let the cardinality of the alphabet become infinite. This approach then provides interesting q-analogues of classical combinatorial quantities since they reduce to permutation statistics when $q\to 1$.

EXERCISE 8. Analyse records in random permutations of a multiset, which corresponds to extracting coefficients $[x_1^{n_1}\cdots x_m^{n_m}]$ in $R'_u(1,z)$ and $\widehat{R}'_u(1,z)$. Derive in this way the fact that the mean number of records in a random permutation of n elements is the harmonic number H_n. [Hint. See [**59**, 1.2.10.18] and [**17**].]

3.3. Paths in graphs, automata, and transfer matrices. A closely related set of applications of regular functions is to problems that are naturally described as paths in digraphs, or equivalently as finite automata. In physics, the corresponding treatment is also called the "transfer matrix method". We start our exposition with the enumeration of paths in graphs that constitutes the most direct introduction to the subject.

3.3.1. *Paths in graphs.* Let G be a directed graph with vertex set $\{1,\ldots,m\}$, where self-loops are allowed and label each edge (a,b) by the (formal or numeric) variable $g_{i,j}$. Consider the matrix $\mathbf{G}$ such that

$$\mathbf{G}_{a,b}=g_{a,b} \text{ if the edge } (a,b)\in G,\qquad \mathbf{G}_{a,b}=0 \text{ otherwise.}$$

Then, from the standard definition of matrix products, the powers $\mathbf{G}^r$ have elements that are path polynomials. More precisely, one has the simple but essential relation,

$$(\mathbf{G})^r_{i,j}=\sum_{w\in\mathcal{P}(i,j;r)} w,$$

where $\mathcal{P}(i,j,r)$ is the set of paths in G that connect i to j and have length r, and a path w is assimilated to the monomial in indeterminates $\{g_{i,j}\}$ that represents multiplicatively the succession of its edges; for instance:

$$(\mathbf{G})^3_{i,j} = \sum_{m_1=i,m_2,m_3,m_4=j} g_{m_1,m_2}g_{m_2,m_3}g_{m_3,m_4},$$

In other words, powers of the matrix associated to a graph "generate" all paths in a graph. One may then treat simultaneously all lengths of paths (and all powers of matrices) by introducing the variable z to record length.

PROPOSITION 8.2. (i) *Let G be a digraph and let $\mathbf{G}$ be the matrix associated to G. The OGF $F^{\langle i,j\rangle}(z)$ of the set of all paths from i to j in a digraph G with z marking length and $g_{a,b}$ marking the occurrence of edge (a,b) is the entry i,j of the matrix $(I-z\mathbf{G})^{-1}$, namely*

$$F^{\langle i,j\rangle}(z) = \left.(I-z\mathbf{G})^{-1}\right|_{i,j} = \frac{\Delta^{\langle i,j\rangle}(z)}{\Delta(z)},$$

where $\Delta(z) = \det(I-z\mathbf{G})$ and $\Delta^{\langle i,j\rangle}(z)$ is the determinant of the minor of index i,j of $I-z\mathbf{G}$.

(ii) *The generating function of nonempty closed paths is given by*

$$\sum_i (F^{\langle i,i\rangle}(z)-1) = -z\frac{\Delta'(z)}{\Delta(z)}.$$

Proof. Part (i) results from the discussion above which implies

$$F^{\langle i,j\rangle}(z) = \sum_{n=0}^{\infty} z^n\,(\mathbf{G}^n)_{i,j} = \left((I-z\mathbf{G})^{-1}\right)_{i,j},$$

and from the cofactor formula of matrix inversion. Part (ii) results from Jacobi's trace formula. Introduce the quantity known as the *zeta function*,

$$\begin{aligned}\zeta(z) \quad &:= \quad \exp\left(\sum_i\sum_{n=1}^{\infty} F_n^{\langle i,i\rangle}\frac{z^n}{n}\right) \quad = \quad \exp\left(\sum_{n=1}^{\infty}\frac{z^n}{n}\mathrm{Tr}\,\mathbf{G}^n\right)\\ &= \quad \exp\left(\mathrm{Tr}\,\log(I-z\mathbf{G})^{-1}\right) \quad = \quad \det(I-z\mathbf{G})^{-1},\end{aligned}$$

where the last line results from the Jacobi trace formula. Thus, $\zeta(z) = \Delta(z)^{-1}$. On the other hand, differentiation combined with the definition of $\zeta(z)$ yields

$$\begin{aligned}z\frac{\zeta'(z)}{\zeta(z)} \quad &= \quad -z\frac{\Delta'(z)}{\Delta(z)}\\ &= \quad \sum_i\sum_{n=1}^{\infty} F_n^{\langle i,i\rangle}z^n,\end{aligned}$$

and Part (ii) follows. □

EXERCISE 9. Can the coefficients of $\Delta(z)$ be related to the polynomials representing self-loops, 2–loops, triangles, quadrangles, etc, in the graph G? [See [**19**]]

EXERCISE 10. Observe that

$$z\frac{\zeta'(z)}{\zeta(z)} = \sum_{n\geq 1} z^n\,\mathrm{Tr}\,\mathbf{G}^n = \sum_\lambda \frac{\lambda z}{1-\lambda z},$$

(the sum is over eigenvalues). Deduce an algorithm that determines the characteristic polynomial of a matrix of dimension m in $\mathcal{O}(m^4)$ arithmetic operations.

[Hint: computing the quantities $\operatorname{Tr} G^j$ for $j = 1, \ldots, m$ requires precisely m matrix multiplications.]

In particular, the number of paths of length n is obtained by applying a substitution $\sigma: g_{a,b} \mapsto 0, 1$ to $(I - z\mathbf{G})^{-1}$ upon coefficient extraction by the $[z^n]$ operation. In a similar vein, it is possible to consider weighted graphs, where the $g_{a,b}$ are assigned real weights; with the weight of a path being defined by the product of its edges weights, one finds that $[z^n](I - z\mathbf{G})^{-1}$ equals the total weight of all paths of length n. If furthermore the assignment is made in such a way that $\sum_b g_{a,b} = 1$, then the matrix $\mathbf{G}$, which is called a stochastic matrix, can be interpreted as the transition matrix of a Markov chain.

Let us assume that nonnegative weights are assigned to the edges of G. If the resulting matrix is irreducible and aperiodic, then Perron-Frobenius theory applies. There exists $\rho = 1/\lambda_1$, with $\lambda_1 > 0$ the dominant eigenvalue of $\mathbf{G}$, and the OGF of weighted paths from i to j has a simple pole at ρ. In that case, it turns out that a random (weighted) path of length n has, asymptotically as $n \to \infty$,

— an average number of edges of type $(i.j)$ that is $\sim \gamma_{i,j} n$, for some nonzero constant $\gamma_{i,j} \in \mathbb{R}$;
— an average number of encounters with vertex i that is $\sim \beta_i n$, for some nonzero constant $\beta_i \in \mathbb{R}$.

In other words, a long random path tends to spend asymptotically a fixed (nonzero) fraction of its time at any given vertex or along any given edge. These observations are the combinatorial counterpart of the elementary theory of finite Markov chains. The treatment of such questions depends on the following lemma.

LEMMA 8.1 (Iteration of Perron-Frobenius matrices). *Set $M(z) = (I - zG)^{-1}$ where G has nonnegative entries, is irreducible, and is aperiodic. Let λ_1 be the dominant eigenvalue of G. Then the "residue" matrix R such that*

$$(I - zG)^{-1} = \frac{R}{1 - z/\lambda_1} + O(1) \qquad (z \to \lambda_1) \tag{9}$$

has entries given by ($\langle x, y\rangle$ represents a scalar product)

$$R_{ij} = \frac{r_i \ell_j}{\langle r, \ell\rangle},$$

where r and ℓ are right and left eigenvectors of G corresponding to the eigenvalue λ_1.

Proof. Scaling z as z/λ_1 reduces the situation to the case of a matrix with dominant eigenvalue equal to 1, so that we assume now $\lambda_1 = 1$. First observe that

$$R = \lim_{n \to \infty} G^n.$$

The limit exists for the following reason: geometrically, G decomposes as $G = P + S$ where P is the projector on the eigenspace generated by the eigenvector r; one has $SP = PS = 0$ and $P^2 = P$, so that $G^n = P^n + S^n$; on the other hand, S has spectral radius < 1; thus $\lim G^n$ exists and it equals P (so that $R \equiv P$ is a projector).

Now, for any vector w, by properties of projections, one has

$$Rw = c(w) r,$$

for some coefficient $c(w)$. Application of this to each of the base vectors e_j (i.e., $e_j = (\delta_{j1}, \ldots, \delta_{jd})$) shows that the matrix R has each of its *columns* proportional to the eigenvector r. A similar reasoning with the transpose G^t of G and the associated residue matrix R^t shows that the matrix R has each of its *rows* proportional to the eigenvector ℓ. In other words, for some constant γ, one has

$$R_{i,j} = \gamma \ell_j r_i.$$

The normalization constant γ is itself finally determined by applying R to r and one finds that $\gamma = 1/\langle \ell, r\rangle$. □

EXERCISE 11. Relate explicitly the edge traversal and node encounter frequencies to the dominant eigenvectors of a graph assumed to be strongly connected and not cyclically layered. Discuss the stationary probabilities of a Markov chain in this context.

EXERCISE 12. What happens when the matrix $\mathbf{G}$ is symmetric (i.e., the graph G is undirected)? Discuss formulæ for reversible Markov chains.

EXAMPLE 6. *Locally constrained words.* Consider a fixed alphabet $\mathcal{A} = \{a_1, \dots, a_m\}$ and a set $\mathcal{F} \subseteq \mathcal{A}^2$ of forbidden transitions between consecutive letters. The set of words over $\mathcal{A}$ with no forbidden transitions is denoted by $\mathcal{L}$ and is called a locally constrained language. Clearly, the words of $\mathcal{L}$ are in bijective correspondence with paths in a graph that is constructed as follows. Set up the complete graph $K_{m\times m}$, where the j vertex of the graph represents the production of the letter a_j. Delete from the complete graph all edges that correspond to forbidden transitions. Then, a word of length $n+1$ has no forbidden transition iff it corresponds to a path of length n in the modified graph. Consequently, the OGF of any locally constrained language is a rational function. Its OGF is given by

$$1 + z(1,1,\dots,1)(I - zT)^{-1}(1,1,\dots,1)^t,$$

where T_{ij} is 0 if $(a_i, a_j) \in \mathcal{F}$ and 1 otherwise. Various specializations, including multivariate GF's and nonuniform letter models are easily treated by this method.

The particular case where $\mathcal{F}$ consists of pairs of equal letters defines Smirnov words [**48**, p. 69] and is amenable to a direct and explicit treatment. Let $W(x_1, \dots, x_m)$ be the multivariate GF of words with the variable x_j marking the number of occurrences of letter a_j and $S(x_1, \dots, x_m)$ be the corresponding GF for Smirnov words. A simple substitution (already discussed in Chapter 3) shows that W and S are related by

$$W(x_1, \dots, x_m) = S\left(\frac{x_1}{1-x_1}, \dots, \frac{x_m}{1-x_m}\right),$$

while $W = (1 - x_1 - \cdots - x_m)^{-1}$. There results that

$$S(x_1, \dots, x_m) = W\left(\frac{x_1}{1+x_1}, \dots, \frac{x_m}{1+x_m}\right) = \left(1 - \sum_{i=1}^{m} \frac{x_j}{1+x_j}\right)^{-1}. \tag{10}$$

In particular, setting $x_j = z$, one gets the univariate OGF

$$S(z) = \frac{1+z}{1-(m-1)z},$$

implying that the number of words of length n is $m(m-1)^{n-1}$ (as it should be). □

Locally constrained languages thus have rational generating functions. The process of proof can then be adapted to the enumeration of many types of objects provided they have a sufficiently "sequential" structure. Carlitz compositions defined below illustrate a recycling of the notion of regular expression to objects that are in fact defined over infinite alphabets.

EXAMPLE 7. *Carlitz compositions.* A Carlitz composition of the integer n is a composition of n such that no two adjacent summands have equal values. Consider first compositions with a bound m on the largest allowable summand. The OGF of Carlitz compositions is directly derived from the GF of Smirnov words by substituting $x_j \mapsto z^j$. In this way, the OGF of Carlitz compositions with maximum summand at most m is found to be

$$C^{[m]}(z) = \left(1 - \sum_{j=1}^{m} \frac{z^j}{1+z^j}\right)^{-1},$$

and the OGF of all Carlitz compositions is obtained by letting m tend to infinity: the preceding GF's converge to

$$C^{[\infty]}(z) = \left(1 - \sum_{j=1}^{\infty} \frac{z^j}{1+z^j}\right)^{-1}. \tag{11}$$

In particular, we get Sequence A003242 of *EIS*[3]:

$$C^{[\infty]}(z) = 1 + z + z^2 + 3z^3 + 4z^4 + 7z^5 + 14z^6 + 23z^7 + 39z^8 + 71z^9 + \cdots .$$

The asymptotic form of the number of Carlitz compositions is then easily found by singularity analysis of meromorphic functions (see Chapter 4). We find

$$C_n^{[\infty]} \sim C \cdot \alpha^n, \quad C \doteq 0.45638, \quad \alpha \doteq 1.75024\,.$$

There, $1/\alpha$ is determined as the smallest positive root of the denominator in (11); see [**57**] for details. In this infinite alphabet example, the OGF is no longer rational but the essential feature of having a dominant polar singularity is preserved. ☐

3.3.2. *Finite automata.* Finite automata are closely related to languages of paths in graphs and to regular expressions.

DEFINITION 8.4 (Finite state automaton). *A finite automaton A over a finite alphabet $\mathcal{A}$ and with vertex set Q, called the set of states, is a digraph whose edges are labelled by letters of the alphabet, that is given together with a designated initial state $q_0 \in Q$ and a designated set of final states $Q_f \subseteq Q$.*

A word w is said to be accepted by the automaton if there exists a path π in the graph connecting the initial state q_0 to one of the final states $q \in Q_f$, so that the succession of labels of the path π corresponds to the sequence of letters composing w. (The path π is then called an accepting path for w.) The set of accepted words is denoted by $\mathcal{L}(A)$.

In all generality, a finite automaton is, by its definition, a *nondeterministic* device: if a word w is accepted, one may not "know" *a priori* which choices of edges to select in order to accept it. A finite automaton is said to be *deterministic* if given any state $q \in Q$ and any letter $x \in \mathcal{A}$, there is at most one edge from vertex q that bears label x. In that case, one decides easily (in linear time) whether a word is accepted by just following edges dictated by the sequence of letters in w.

[3]The *EIS* designates Sloane's On-Line Encyclopedia of Integer Sequences [**85**]; see [**86**] for an earlier printed version.

The main structural result of the theory is that there is complete equivalence between three descriptive models: regular expressions, deterministic finite automata, and nondeterministic finite automata. The corresponding theorems are due to Kleene (the equivalence between regular expression and nondeterministic finite automata) and to Rabin and Scott (the equivalence between nondeterministic and deterministic automata). Thus, finite automata whether deterministic or not accept ("recognize") the class of all regular languages.

EXERCISE 13. *Any language definable by a regular expression is accepted by a (non-deterministic) finite automaton.* [Hint: Perform an inductive construction of an automaton based on separate constructions for unions, products, stars.]

EXERCISE 14. Kleene's theorem: *Any language accepted by a finite automaton is definable by a regular expression.* [Hint: Construct (inductively on increasing values of k) matrices of regular expressions where $R_{i,j}^{(k)}$ is a regular expression that describes the set of paths from state i to state j constrained never to pass through a state of index $\geq k$.]

EXERCISE 15. Rabin and Scott's theorem: *Given any finite automaton A, there exists an equivalent deterministic automation B, i.e., $\mathcal{L}(A) = \mathcal{L}(B)$.* [Hint: the states of B are all the subsets of the states of A and the transitions are determined so that B emulates all possible computations of A.]

EXERCISE 16. Regular languages are closed under intersection and complementation.

PROPOSITION 8.3 (Finite state automata counting). *Any language accepted by a finite automaton or described by a regular expression has a rational generating function. If the language is specified by an automaton $A = \langle Q, Q_f, q_0 \rangle$, that is deterministic, then the corresponding ordinary generating function $L_0(z)$ is defined as the component $L_0(z)$ of the linear system of equations*

$$L_j(z) = \phi_j + z \sum_{a \in \mathcal{A}} L_{\tau(q_j,a)}(z),$$

where ϕ_j equals 1 if $q_j \in Q_f$ and 0 otherwise, and where $\tau(q_j, a)$ is the state reachable from state q_j when the letter a is read.

Proof. By the fundamental equivalence of models, one may freely assume the automaton to be deterministic. The quantity L_j is nothing but the OGF of the language obtained by changing the initial state of the automaton to q_j. Each equation expresses the fact that a word accepted starting from q_j may be the empty word (if q_j is final) or, else, it must consist of a letter a followed by a continuation that is itself accepted when the automaton is started from the "next" state, that is, the state of index $\tau(q_j, a)$.

Equivalently, one may reduce the proof to the enumeration of paths in graphs as detailed above. □

Given a problem that is described by a regular language, it is then a matter of choice to enumerate it using a specification by a regular expression (provided it is unambiguous) or by a finite automaton (provided it is deterministic). Each problem usually has a more natural presentation. From the structural standpoint, it should however be kept in mind that the conversion between a regular expression and an equivalent nondeterministic automata involves a moderate (at worst polynomial) blow-up in size, while the transformation of a nondeterministic automaton to an equivalent deterministic one may involve an exponential increase in the number of states. (The construction underlying the equivalence theorem of Rabin and Scott involves the power-set of the set of states Q of the nondeterministic automaton.) Note that the direct construction of the generating functions associated to finite automata has been already encountered in Chapter 1 when discussing particular cases of languages containing or excluding a fixed pattern.

As the proof of the proposition shows, the OGF of the language defined by a deterministic finite automaton involves a quasi-inverse $(1 - zT)^{-1}$ where the transition matrix T is a direct encoding of the automaton's transitions. Corollary 8.2 and Lemma 8.1 were precisely custom-tailored for this situation. Consequently, quantifying probabilistic phenomena associated to finite automata is invariably reducible to finding eigenvectors of matrices.

EXAMPLE 8. *Words with excluded patterns.* Fix a pattern $u = u_1 u_2 \cdots u_m$. Let E_u be the set of words that do not contain the word u as a factor, and F_u the set of words that do not contain u as a subsequence. The language E_u is defined by the (ambiguous) regular expression $\mathcal{A}^\star \setminus (\mathcal{A}^\star u \mathcal{A}^\star)$. A deterministic finite automaton for $\mathcal{A}^\star u \mathcal{A}^\star$ is readily constructed from u, from which a deterministic automaton results for the complement E_u. (It suffices to exchange final and nonfinal states.) Thus, the generating function of E_u is rational.

A similar argument, starting from the (ambiguous) regular expression that describes F_u, namely

$$\mathcal{A}^\star u_1 \mathcal{A}^\star\, u_2 \mathcal{A}^\star \cdots \mathcal{A}^\star u_m \mathcal{A}^\star,$$

also shows (via the construction of the finite automaton) that the OGF of F_u is rational. □

EXERCISE 17. Extend the previous argument to prove the rationality of the OGF of words that do not contain any pattern from a finite set S, where pattern is taken either in the sense of a factor or a subsequence.

EXERCISE 18. Calculate explicitly the generating function of F_u for an arbitrary u in the case of a binary and of a ternary alphabet.

Determine the mean number of occurrences of u as subsequence in a random text of size n. Discuss the asymptotic form of the variance of the number of such occurrences.

EXAMPLE 9. *Variable-length codes.* A finite set of words $\mathcal{S} \subset \mathcal{A}^\star$ is a ("variable-length") *code* [**11**] if each word in $\mathcal{A}^\star$ admit at most one decomposition as a concatenation of words, each of which belonging to $\mathcal{S}$. Obviously, such a code can be used for encoding informations, adjusting formats of data to a particular transmission channel, etc.

It is not clear *a priori* how to decide whether any given set S is a code. Let $S(z) = \sum_{w \in \mathcal{S}} z^{|w|}$ be the OGF of $\mathcal{S}$ (a polynomial). Then $\mathcal{S}$ is a code iff

$$\frac{1}{1 - S(z)} = L(z),$$

where $L(z)$ is the GF of all words that admit a decomposition as concatenations of words in $\mathcal{S}$. Now, Aho and Corasick have established in [**3**] that there exists a deterministic automaton of size linear in the total size of S (the total number of letters) that recognizes $S^\star$ (furthermore this automaton can be constructed in linear time). Thus, a system determining $L(z)$ is found in linear time.

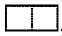
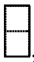
From the combinatorial considerations above, there results a decision procedure for codicity (based on linear algebra) that is of polynomial time complexity. □

EXERCISE 19. A finite state automaton is said to be unambiguous if the set of accepting paths of any given words comprises at most one element. Prove that the translation into generating function as described above also applies to such automata, even if they are nondeterministic.

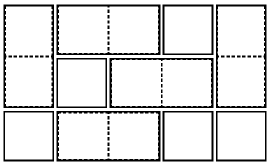
3.3.3. *Transfer matrix methods.* The transfer matrix method constitutes a variant of the modelling by deterministic automata and the encoding of combinatorial problems by regular languages. The very general statement of Theorem 8.6 applies here with full strength. Here, we shall simply illustrate the situation by an example inspired by the insightful exposition of domino tilings and generating functions in the book of Graham, Knuth, and Patashnik [**50**].

EXAMPLE 10. *Monomer-dimer tilings of a rectangle.* Suppose one is given pieces that may be one of the three forms: monomers (m) that are 1×1 squares, and dimers that are dominoes, either vertically (v) oriented 1×2, or horizontally (h) oriented 2×1. In how many ways can an $n \times 3$ rectangle be covered completely and without overlap ('tiled') by such pieces?

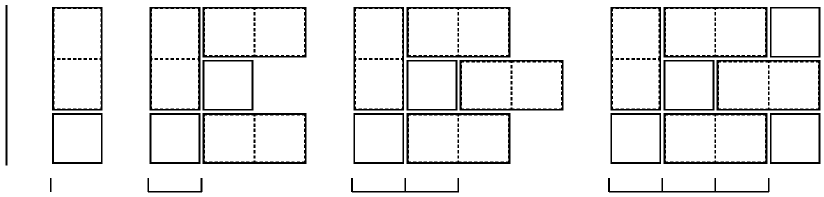
The pieces are thus of the following types,

$m =$ , $h =$ , $v =$ ,

and here is a particular tiling of a 5×3 rectangle:



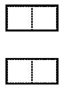
In order to approach this counting problem, one defines a class $\mathcal{C}$ of combinatorial objects called configurations. A configuration relative to an $n \times k$ rectangle is a partial tiling, such that all the first $n-1$ columns are entirely covered by dominoes while between zero and three unit cells of the last column are covered. Here are for instance, configurations corresponding to the example above.



These diagrams suggest the way configurations can be built by successive addition of dominoes. Starting with the empty rectangle 0×3, one adds at each stage a collection of at most three dominoes in such a way that there is no overlap. This creates a configuration where, like in the example above, the dominoes may not be aligned in a flush-right manner. Continue to add successively dominoes whose left border is at abscissa $1, 2, 3$, etc, in a way that creates no internal "holes".

Depending on the state of filling of their last column, configuration can thus be classified into 8 classes that we may index in binary as $\mathcal{C}_{000}, \ldots, \mathcal{C}_{111}$. For instance $\mathcal{C}_{001}$

represent configurations such that the first two cells (from top to bottom, by convention) are free, while the third one is occupied. Then, a set of rules describes the new type of configuration obtained, when the sweep line is moved one position to the right and dominoes are added. For instance, we have

$\mathcal{C}_{010} \quad \odot$  $\quad \Longrightarrow \quad \mathcal{C}_{101}.$

In this way, one can set up a grammar (resembling a deterministic finite automaton) that expresses all the possible constructions of longer rectangles from shorter ones according to the last layer added. The grammar comprises productions like

$$\begin{aligned} C_{000} \quad = \quad & \epsilon + \underline{mmm}C_{000} + \underline{mv}C_{000} + \underline{vm}C_{000} \\ & + \underline{\cdot mm}C_{100} + \underline{m\cdot m}C_{010} + \underline{mm\cdot}C_{001} + \underline{v\cdot}C_{001} + \underline{\cdot v}C_{100} \\ & + \underline{m\cdot\cdot}C_{011} + \underline{\cdot m\cdot}C_{101} + \underline{\cdot\cdot m}C_{110} + \underline{\cdot\cdot\cdot}C_{111}\ . \end{aligned}$$

In this grammar, a "letter" like $\underline{mv}$ represent the addition of dominoes, in top to bottom order, of types m, v respectively; the letter $\underline{m\cdot m}$ means adding two m-dominoes on the top and on the bottom, etc.

The grammar transforms into a linear system of equations with polynomial coefficients. The substitution $m \mapsto z$, $h, v \mapsto z^2$ then gives the generating functions of configurations with z marking the area covered:

$$C_{000}(z) = \frac{(1-2z^3-z^6)(1+z^3-z^6)}{(1+z^3)(1-5z^3-9z^6+9z^9+z^{12}-z^{15})}..$$

In particular, the coefficient $[z^{3n}]C_{000}(z)$ is the number of tilings of an $n \times 3$ rectangle:

$$C_{000}(z) = 1 + 3z^3 + 22z^6 + 131z^9 + 823z^{12} + 5096z^{15} + \cdots .$$

The sequence grows like $c\,\alpha^n$ (for $n \equiv 0 \pmod 3$) where $\alpha \doteq 1.83828$ (α is the cube root of an algebraic number of degree 5). (See [**20**] for a computer algebra session.) On average, for large n, there is a fixed proportion of monomers and the distribution of monomers in a random tiling of a large rectangle is asymptotically normally distributed, as results from the developments of Chapter 9. $\square$

As is typical of the tiling example, one seeks to enumerate a "special" set of configurations $\mathcal{C}_f$. (In the example above, this is $\mathcal{C}_{000}$ representing complete rectangle coverings.) One determines an extended set of configurations $\mathcal{C}$ (the partial coverings, in the example) such that: (i) $\mathcal{C}$ is partitioned into finitely many classes; (ii) there is a finite set of "actions" that operate on the classes; (iii) size is affected in a well-defined additive way by the actions. The similarity with finite automata is apparent: classes play the rôle of states and actions the rôle of letters.

EXERCISE 20. For any fixed width w, the OGF of monomer-dimer coverings of an $n \times w$ rectangle is rational.

The OGF of Hamiltonian tours on an $n \times w$ rectangle is rational (one is allowed to move from any cell to any other vertically or horizontally adjacent cell). The same holds for king's tours and knight's tours.

EXERCISE 21. The OGF of trees (binary, general) of bounded width is a rational function.

EXERCISE 22. Find combinatorial interpretations of

$$F^{[w]}(z) := \sum_{n=0}^{\infty} (F_n)^w \, z^n,$$

for any fixed integer w, where the numbers F_n are the Fibonacci numbers. [Hint: see [**50**].]

EXERCISE 23. Given a fixed digraph G assumed to be strongly connected, and a designated start vertex, one travels at random, moving at each time to any neighbour of the current vertex chosen with equal likelihood. Show that the expectation of the time to visit all the vertices is a rational number that is effectively (though perhaps not efficiently!) computable.

Often, the method of transfer matrices is used to approximate a hard combinatorial problem that is not known to decompose, the approximation being by means of a family of models of increasing "widths". For instance, the enumeration of the number T_n of tilings of an $n \times n$ square by monomers and dimers remains a famous unsolved problem of statistical physics. Here, transfer matrix methods may be used to solve the $n \times w$ version of the monomer–dimer coverings, in principle at least, for any fixed width w. (The "diagonal" sequence of the $n \times w$ rectangular models corresponds to the square model.) It has been at least determined by computer search that the diagonal sequence T_n starts as (this is sequence A028420 in Sloane's *EIS* [**86**]):

$$1,\ 7,\ 131,\ 10012,\ 2810694,\ 2989126727,\ 11945257052321, \ldots .$$

From this and other numerical data, one estimates numerically that $(T_n)^{1/n^2} \to 1.94021\ldots$, but no expression for the constant is known to exist[4]. The difficulty of coping with the finite-width models is that their complexity (as measured , e.g., by the number of states) blows up exponentially with w—such models are best treated by computer algebra; see [**102**]—and no law allowing to take a diagonal is visible. However, the finite width models have the merit of providing at least provable upper and lower bounds on the exponential growth rate of the hard "diagonal problem".

3.4. Permutations and local constraints. In this subsection, we examine problems whose origin lies in nineteenth century recreational mathematics. For instance, the *ménage* problem solved and popularized by Édouard Lucas in 1891, see [**26**], has the following quaint formulation: *What is the number of possible ways one can arrange n married couples (*'ménages'*) around a table in such a way that men and women alternate, but no woman sits next to her husband?*

The ménage problem is equivalent to a permutation enumeration problem. Sit first conventionally the men at places numbered $0, \ldots, n-1$, and let σ_i be the position at the right of which the ith wife is placed. Then, a ménage placement imposes the condition $\sigma_i \neq i$ and $\sigma_i \neq i+1$ for each i. We consider here a linearly arranged table (see remarks at the end for the other classical formulation that considers a round table), so that the

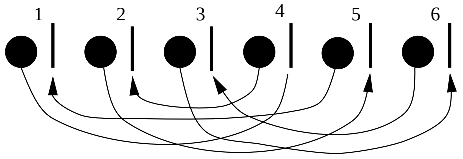
[4] In contrast, for coverings by dimers only, a strong algebraic structure is available and the number of covers $\widehat{T}_n$ satisfies

$$\lim_{2n \to +\infty} \left(\widehat{T}_{2n}\right)^{1/(2n)^2} = \exp\left(\frac{1}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2}\right) = e^{G/\pi} \doteq 1.33851\ldots,$$

where G is Catalan's constant; see [**76**] for context on this famous result.

condition $\sigma_i \neq i+1$ becomes vacuous when $i = n$. Here is a ménage placement for $n = 6$ corresponding to the permutation

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 2 & 1 & 3 \end{bmatrix}$$



Clearly, this is a generalization of the derangement problem (for which the weaker condition $\sigma_i \neq i$ is imposed), where the cycle decomposition of permutations suffices to provide a direct solution (see Chapter 2).

Given a permutation $\sigma = \sigma_1 \cdots \sigma_n$, any quantity $\sigma_i - i$ is called an *exceedance* of σ. Let Ω be a finite set of integers that we assume to be nonnegative. Then a permutation is said to be Ω-avoiding if none of its exceedances lies in Ω. The counting problem, as we now demonstrate, provides an interesting case of application of the transfer matrix method.

The set Ω being fixed, consider first for all j the class of augmented permutations $\mathcal{P}_{n,j}$ that are permutations of size n such that j of the positions are distinguished and the corresponding exceedances lie in Ω, the remaining positions having arbitrary values (but with the permutation property being satisfied!). Loosely speaking, the objects in $\mathcal{P}_{n,j}$ can be regarded as permutations with "at least" j exceedances in Ω. For instance, with $\Omega = \{1\}$ and

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 4 & 8 & 6 & 7 & 1 & 5 & 9 \end{pmatrix},$$

there are 5 exceedances that lie in Ω (at positions $1, 2, 3, 5, 6$) and with 3 of these distinguished (say by enclosing them in a box), one obtains an element counted by $\mathcal{P}_{9,3}$ like

$$2\,\boxed{3}\,\boxed{4}\,8\,6\,\boxed{7}\,1\,5\,9.$$

Let $P_{n,j}$ be the cardinality of $\mathcal{P}_{n,j}$. We claim that the number $Q_n = Q_n^{\Omega}$ of Ω-avoiding permutations of size n satisfies

$$Q_n = \sum_{j=0}^{n} (-1)^j P_{n,j}. \tag{12}$$

Equation (12) is typically an *inclusion-exclusion* relation. To prove it formally, define the number $R_{n,k}$ of permutations that have exactly k exceedances in Ω and the generating polynomials

$$P_n(w) = \sum_j P_{n,j} w^j, \qquad R_n(w) = \sum_k R_{n,k} w^k.$$

The GF's are related by

$$P_n(w) = R_n(w+1) \qquad \text{or} \qquad R_n(w) = P_n(w-1)..$$

(The relation $P_n(w) = R_n(w+1)$ simply expresses symbolically the fact that each Ω-exceedance in $\mathcal{R}$ may or may not be taken in when composing an element of $\mathcal{P}$.) In particular, we have $P_n(-1) = R_n(0) = R_{n,0} = Q_n$ as was to be proved.

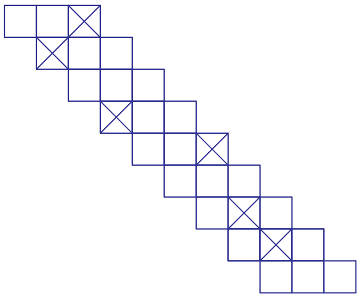


FIGURE 5. A graphical rendering of the legal template **20?02?11?** relative to $\Omega = \{0, 1, 2\}$.

The preceding discussion shows that everything relies on the enumeration $P_{n,j}$ of permutations with distinguished exceedances in Ω. Introduce the alphabet $\mathcal{A} = \Omega \cup \{\text{'?'}\}$, where the symbol '?' is called the 'don't-care symbol'. A word on $\mathcal{A}$, an instance with $\Omega = \{0, 1, 2\}$ being 20?02?11?, is called a *template*. To an augmented permutation, one associates a template as follows: each exceedance that is not distinguished is represented by a don't care symbol; each distinguished exceedance (thereby an exceedance with value in Ω) is represented by its value. A template is said to be legal if it arises from an augmented permutation. For instance a template $2\,1\cdots$ cannot be legal since the corresponding constraints, namely $\sigma_1 - 1 = 2$, $\sigma_2 - 2 = 1$, are incompatible with the permutation structure (one should have $\sigma_1 = \sigma_2 = 3$). In contrast, the template 20?02?11? is seen to be legal. Figure 5 is a graphical rendering; there, letters of templates are represented by dominoes, with a cross at the position of a numeric value in Ω, and with the domino being blank in the case of a don't-care symbol.

Let $T_{n,j}$ be the set of legal templates relative to Ω that have length n and comprise j don't care symbols. Any such legal template is associated to exactly $j!$ permutations, since $n-j$ position-value pairs are fixed in the permutation, while the j remaining positions and values can be taken arbitrarily. There results that

$$P_{n,n-j} = j!\,T_{n,j} \qquad \text{and} \qquad Q_n = \sum_{j=0}^{n} (-1)^{n-j} j!\,T_{n,j}, \tag{13}$$

by (12). Thus, the enumeration of avoiding permutations rests entirely on the enumeration of legal templates.

The enumeration of legal templates is finally effected by means of a transfer matrix method, or equivalently, by a finite automaton. If a template $\tau = \tau_1 \cdots \tau_n$ is legal, then the following condition is met,

$$\tau_j + j \neq \tau_i + i, \tag{14}$$

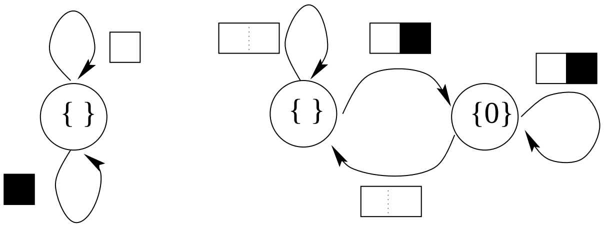
for all pairs (i, j) such that $i < j$ and neither of τ_i, τ_j is the don't-care symbol. (There are additional conditions to characterize templates fully, but these only concern a few letters at the end of templates and we may ignore them in this discussion.) In other words, a τ_i with a numerical value preempts the value $\tau_i + i$. Figure 5 exemplifies the situation in the case $\Omega = \{0, 1, 2\}$. The dominoes are shifted one position each time (since it is the value of $\sigma - i$ that is represented) and the compatibility constraint (14) is that no two crosses should be vertically aligned. More precisely the constraints (14) are recognized by a deterministic finite automaton whose states are indexed by subsets of $\{0, \ldots, b-1\}$ where the "span" b

is defined as $b = \max_{\omega\in\Omega} \omega$. The initial state is the one associated with the empty set (no constraint is present initially), the transitions are of the form

$$\begin{cases} (q_S, j) \ \mapsto \ q_{S'} & \text{where } S' = ((S-1) \cup \{j-1\}) \cap \{0, \ldots, b-1\}, \ j \neq \text{‘?’} \\ (q_S, ?) \ \mapsto \ q_{S'} & \text{where } S' = (S-1) \cap \{0, \ldots, b-1\}; \end{cases}$$

the final state is equal to the initial state (this translates the fact that no domino can protrude from the right, and is implied by the linear character of the ménage problem under consideration). In essence, the automaton only needs a finite memory since the dominoes slide along the diagonal and, accordingly, constraints older than the span can be forgotten. Notice that the complexity of the automaton, as measured by its number of states, is 2^b.

Here are the automata corresponding to $\Omega = \{0\}$ (derangements) and to $\Omega = \{0, 1\}$ (ménages).



For the ménage problem, there are two states depending on whether or not the currently examined value has been preempted at the preceding step.

From the automaton construction, the bivariate GF $T^\Omega(z, u)$ of legal templates, with u marking the position of don't care symbols, is a rational function that can be determined in an automatic fashion from Ω. For the derangement and ménage problems, one finds

$$T^{\{0\}}(z, u) = \frac{1}{1 - z(1+u)}, \qquad T^{\{0,1\}}(z, u) = \frac{1 - z}{1 - z(2+u) + z^2}.$$

In general, this gives access to the OGF of the corresponding permutations. Consider the partial expansion of $T^\Omega(z, u)$ with respect to u, taken under the form

$$T^\Omega(z, u) = \sum_r \frac{c_r(z)}{1 - u u_r(z)}, \tag{15}$$

assuming for convenience only simple poles. There the sum is finite and it involves algebraic functions c_j and u_j of the variable z. Finally, the OGF of Ω-avoiding permutations is obtained from T^Ω by the transformation

$$z^n u^k \quad \mapsto \quad (-z)^n k!,$$

which is the transcription of (13). Define the (divergent) OGF of all permutations,

$$F(y) = \sum_{n=0}^{\infty} n!\, y^n = {}_2F_0[1, 1; y],$$

in the terminology of hypergeometric functions. Then, by the remarks above and (15), we find

$$Q^\Omega(z) = \sum_r c_r(-z) F(-u_j(-z)).$$

In other words, *the OGF of* Ω*-avoiding permutations is a combination of compositions of the OGF of the factorial series with algebraic functions.*

The expressions simplify much in the case of ménages and derangements where the denominators of T are of degree 1 in u. One has

$$Q^{\{0\}}(z) = \frac{1}{1+z}F(\frac{z}{1+z}) = 1 + z^2 + 2z^3 + 9z^4 + 44z^5 + 265z^6 + 1854z^7 + \cdots,$$

for derangements, whence a new derivation of the known formula,

$$Q_n^{\{0\}} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)!.$$

Similarly, for (linear) ménage placements, one finds

$$Q^{\{0,1\}}(z) = \frac{1}{1+z}F(\frac{z}{(1+z)^2}) = 1 + z^3 + 3z^4 + 16z^5 + 96z^6 + 675z^7 + \cdots,$$

which is *EIS*–A000271 and corresponds to the formula

$$\mathbb{Q}_n^{\{0,1\}} = \sum_{k=0}^{n} (-1)^k \binom{2n-k}{k} (n-k)!.$$

Finally, the same techniques adapts to constraints that "wrap around", that is, constraints taken modulo n. (This corresponds to a round table in the ménage problem.) In that case, what should be considered is the loops in the automaton recognizing templates (see also the previous discussion of the zeta function of graphs). One finds in this way the OGF of the circular (i.e., classical) ménage problem to be (*EIS*–A000179)

$$\widehat{Q}^{\{0,1\}}(z) = \frac{1-z}{1+z}F(\frac{z}{(1+z)^2}) + 2z = 1 + z + z^3 + 2z^4 + 13z^5 + 80z^6 + 579z^7 + \cdots,$$

which yields the classical solution of the (circular) ménage problem,

$$\widehat{Q}_n^{\{0,1\}} = \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k} \binom{2n-k}{k} (n-k)!,$$

a formula that is due to Touchard; see [**26**, p. 185] for pointers to the vast classical literature on the subject. The algebraic part of the treatment above is close to the inspiring discussion offered in Stanley's book [**87**]. An application to robustness of interconnections in random graphs is presented in [**38**].

For asymptotic analysis purposes, the following general property proves useful: *Let F be the OGF of factorial numbers and assume that $y(z)$ is analytic at the origin where it satisfies $y(z) = z - \lambda z^2 + O(z^3)$; then it is true that*

$$[z^n]F(y(z)) \sim [z^n]F(z(1-\lambda z)) \sim n!e^{-\lambda}. \tag{16}$$

(The proof results from simple manipulations of divergent series in the style of [**8**].) This gives at sight the estimates

$$Q_n^{\{0\}} \sim ne^{-1}, \qquad Q_n^{\{0,1\}} \sim ne^{-2}.$$

More generally, for any set Ω containing λ elements, one has

$$Q_n^{\{\Omega\}} \sim ne^{-\lambda}.$$

Furthermore, the number $R_{n,k}^{\Omega}$ of permutations having exactly k occurrences (k fixed) of an exceedance in Ω is asymptotic to

$$Q_n^{\{\Omega\}} \sim ne^{-\lambda}\frac{\lambda^k}{k!}.$$

In other words, the rare event that an exceedance belongs to Ω obeys of Poisson distribution with $\lambda = |\Omega|$. These last two results are established by means of probabilistic techniques in the book [**7**, Sec. 4.3]. The relation (16) points to a way of arriving at such estimates by purely analytic-combinatorial techniques.

EXERCISE 24. Given a permutation $\sigma = \sigma_1 \cdots \sigma_n$, a *succession gap* is defined as any difference $\sigma_{i+1} - \sigma_i$. Discuss the counting of permutations whose succession gaps are constrained to lie outside of a finite set Ω.

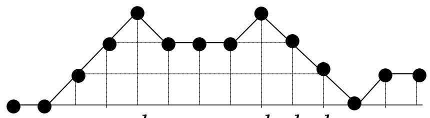
In how many ways can a kangaroo pass through all points of the integer interval $[1, n]$ starting at 1 and ending at n while making hops that belong to $\{-2, -1, 1, 2\}$?

3.5. Lattice paths and walks on the line. In this section, we consider *lattice paths* that are fundamental objects of combinatorics. Indeed, they relate to trees, permutations, and set partitions, to name a few. They also correspond to walks on the integer half-line and as such they relate to classical random walks and to birth-and-death processes of probability theory. The lattice paths discussed here have steps that correspond to movements either immediately to the left or to the right. Combinatorially, such paths are the limit of paths of bounded height, themselves definable as nested sequences. As a consequence, the OGF's obtained are of the continued fraction type.

DEFINITION 8.5 (Lattice path). *A (lattice)* path $\upsilon = (U_0, U_1, \ldots, U_n)$ *is a sequence of points in the lattice* $\mathbb{N} \times \mathbb{N}$ *such that if* $U_j = (x_j, y_j)$*, then* $x_j = j$ *and* $|y_{j+1} - y_j| \le 1$*. An edge* $\langle U_j, U_{j+1} \rangle$ *is called an* ascent *($\underline{a}$) if* $y_{j+1} - y_j = +1$*, a* descent *($\underline{b}$) if* $y_{j+1} - y_j = -1$*, and a* level step *($\underline{c}$) if* $y_{j+1} - y_j = 0$.

The quantity n *is the* length *of the path,* $o(\upsilon) := y_0$ *is the* initial altitude*,* $h(\upsilon) := y_n$ *is the* final altitude*. The extremal quantities* $\sup\{\upsilon\} := \max_j y_j$ *and* $\inf\{\upsilon\} := \min_j y_j$ *are called the* height *and* depth *of the path.*

It is assumed that paths are normalized by the condition $x_0 = 0$. With this normalization, a path of length n is encoded by a word with a, b, c representing ascents, descents, and level steps, respectively. What we call the *standard encoding* is such a word in which each step a, b, c is (redundantly) subscripted by the value of the y-coordinate of its associated point. For instance,



$$w = c_0\, a_0\, a_1\, a_2\, b_3\, c_2\, c_2\, a_2\, b_3\, b_2\, b_1\, a_0\, c_1$$

encodes a path that connects the initial point $(0, 0)$ to the point $(13, 1)$.

Let $\mathcal{H}$ be the set of all lattice paths. Given a geometric condition (Q), it is then possible to associate to it a "language" $\mathcal{H}[Q]$ that comprises the collection of all path encodings satisfying the condition Q. This language can be viewed either as a set or as a formal sum,

$$H[Q] = \sum_{\{w \mid Q\}} w,$$

in which case it becomes the generating function in infinitely many indeterminates of the corresponding condition.

The general subclass of paths of interest in this subsection is defined by arbitrary combinations of flooring (m), ceiling (h), as well as fixing initial (k) and final (l) altitudes:

$$H_{k,l}^{[\ge m, < h]} = \{w \in H : o(w) = k,\ h(w) = l,\ \inf\{w\} \ge m,\ \sup\{w\} < h\}.$$

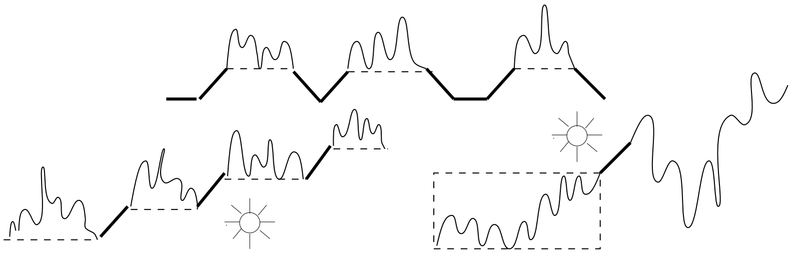


FIGURE 6. The three major decompositions of lattice paths: the arch decomposition (top), the last passages decomposition (bottom left), and the first passage decomposition (bottom right).

We also need the specializations, $H_{k,l}^{[<h]} = H_{k,l}^{[\geq 0,<h]}$, $H_{k,l}^{[\geq m]} = H_{k,l}^{[\geq m,<\infty]}$, $H_{k,l} = H_{k,l}^{[\geq 0,<\infty]}$. Three simple combinatorial decompositions of paths then suffice to derive all the basic formulæ.

Arch decomposition: An excursion from and to level 0 consists of a sequence of "arches", each made of either a c_0 or a $a_0\mathcal{H}_{1,1}^{[\geq 1]}b_1$, so that

$$\mathcal{H}_{0,0} = \left(c_0 \cup a_0\mathcal{H}_{1,1}^{[\geq 1]}b_1\right)^\star,$$

which relativizes to height $< h$.

Last passages decomposition. Recording the times at which each level $0,\ldots,k$ is last traversed gives

$$\mathcal{H}_{0,k} = \mathcal{H}_{0,0}^{[\geq 0]}a_0\mathcal{H}_{1,1}^{[\geq 1]}a_1\cdots a_{k-1}\mathcal{H}_{k,k}^{[\geq k]} \tag{17}$$

First passage decomposition. The quantities $H_{k,l}$ with $k \leq l$ are implicitly determined by the first passage through k in a path connecting level 0 to l, so that

$$\mathcal{H}_{0,l} \quad = \quad \mathcal{H}_{0,k-1}^{[<k]}a_{k-1}\mathcal{H}_{k,l} \quad (k \leq l), \tag{18}$$

A dual decomposition holds when $k \geq l$.

The basic results express the generating functions in terms of a fundamental continued fraction and its associated convergent polynomials. They involve the "numerator" and "denominator" polynomials, denoted by P_h and Q_h that are defined as solutions to the second order (or "three-term") recurrence equation

$$Y_{h+1} = (1-c_h)Y_h - a_{h-1}b_hY_{h-1},\ h \geq 1, \tag{19}$$

together with the initial conditions $(P_{-1},Q_{-1}) = (1,0)$, $(P_0,Q_0) = (0,1)$, and with the convention $a_{-1}b_0 = 1$.

PROPOSITION 8.4 (Path continued fractions [**35**]). (i) *The generating function $H_{0,0}$ of all basic excursions is represented by the fundamental continued fraction:*

$$H_{0,0} \quad = \quad \cfrac{1}{1-c_0-\cfrac{a_0b_1}{1-c_1-\cfrac{a_1b_2}{1-c_2-\cfrac{a_2b_3}{\ddots}}}}. \tag{20}$$

(ii) *The generating function of ceiled excursions* $H_{0,0}^{[<h]}$ *is given by a convergent of the fundamental fraction:*

$$H_{0,0}^{[<h]} = \cfrac{1}{1-c_0-\cfrac{a_0b_1}{1-c_1-\cfrac{a_1b_2}{1-c_2-\cfrac{a_2b_3}{\cfrac{\ddots}{1-c_{h-1}}}}}} \tag{21}$$

$$= \frac{P_h}{Q_h}. \tag{22}$$

(iii) *The generating function of floored excursions is given by the truncation of the fundamental fraction:*

$$H_{h,h}^{[\geq h]} = \cfrac{1}{1-c_h-\cfrac{a_hb_{h+1}}{1-c_{h+1}-\cfrac{a_{h+1}b_{h+2}}{1-c_{h+2}-\cfrac{a_{h+2}b_{h+3}}{\ddots}}}} \tag{23}$$

$$= \frac{1}{a_{h-1}b_h}\,\frac{Q_hH_{0,0}-P_h}{Q_{h-1}H_{0,0}-P_{h-1}}, \tag{24}$$

Proof. Repeated use of the arch decomposition provides a form of $H_{0,0}^{[<h]}$ with nested quasi-inverses $(1-f)^{-1}$ that is the finite fraction representation (21). The continued fraction representation for basic paths (namely $H_{0,0}$) is then obtained by letting $h \to \infty$ in (21). Finally, the continued fraction form (23) for ceiled excursions is nothing but the fundamental form (20), when the indices are shifted. The three continued fraction expressions (20), (21), (23) are thence established.

Finding explicit expressions for the fractions $H_{0,0}^{[<h]}$ and $H_{h,h}^{[\geq h]}$ next requires determining the polynomials that appear in the convergents of the basic fraction (20). By definition, the convergent polynomials P_h and Q_h are the numerator and denominator of the fraction $H_{0,0}^{[<h]}$. For the computation of $H_{0,0}^{[<h]}$ and P_h, Q_h, one classically introduces the linear fractional transformations

$$g_j(y) = \frac{1}{1-c_j-a_jb_{j+1}y},$$

so that

$$H_{0,0}^{[<h]} = g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(0) \text{ and } H_{0,0} = g_0 \circ g_1 \circ g_2 \circ \cdots\ ,. \tag{25}$$

Now, linear fractional transformations are representable by 2×2-matrices

$$\frac{ay+b}{cy+d} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{26}$$

	Type	*Spec.*	*Formula*
1.	Excursion	$H_{0,0}$	$\dfrac{1}{1-c_0} + \dfrac{a_0 b_1}{1-c_1} + \cdots$
2.	Ceiled excursions	$H_{0,0}^{[<h]}$	$\dfrac{P_h}{Q_h}$
3.	Floored excursions	$H_{h,h}^{[\geq h]}$	$\dfrac{1}{a_{h-1}b_h}\dfrac{Q_h H_{0,0} - P_h}{Q_{h-1}H_{0,0} - P_{h-1}}$
4.	Transitions from 0	$H_{0,l}$	$\dfrac{1}{B_l}\left(Q_l H_{0,0} - P_l\right)$
5.	Transitions to 0	$H_{k,0}$	$\dfrac{1}{A_k}(Q_k H_{0,0} - P_k)$
6.	Upcrossings from 0	$H_{0,h-1}^{[<h]}$	$\dfrac{A_{h-1}}{Q_h}$
7.	Downcrossings to 0	$H_{h-1,0}^{[<h]}$	$\dfrac{B_{h-1}}{Q_h}$
8.	Transitions ($k \leq l$)	$H_{k,l}$	$\dfrac{1}{A_k B_l} Q_k \left(Q_l H_{0,0} - P_l\right)$
9.	Transitions ($k \geq l$)	$H_{k,l}$	$\dfrac{1}{A_k B_l} Q_l \left(Q_k H_{0,0} - P_k\right)$
10.	Upward excursions	$H_{m,m}^{[\geq m,<h]}$	$\dfrac{1}{a_{m-1}b_m}\dfrac{D_{m,h}}{D_{m-1,h}}$
11.	Downward excursions	$H_{l,l}^{[<l+1]}$	$\dfrac{Q_l}{Q_{l+1}}$
12.	Transitions in strip ($k \leq l$)	$H_{k,l}^{[\geq m,<h]}$	$\dfrac{1}{A_k B_l}\dfrac{D_{m-1,k}D_{l,h}}{D_{m-1,h}}$
13.	Transitions in strip ($l \leq k$)	$H_{k,l}^{[\geq m,<h]}$	$\dfrac{1}{A_k B_l}\dfrac{D_{m-1,l}D_{k,h}}{D_{m-1,h}}$

TABLE 1. Generating functions associated to some major path conditions. The basic continued fraction is $H_{0,0}$ in Entry 1, with convergent polynomials P_h, Q_h. Abbreviations used are: $A_m = a_0 \cdots a_{m-1}$, $B_m = b_1 \cdots b_m$, and $D_{i,j} = Q_i P_j - P_i Q_j$.

in such a way that composition corresponds to matrix product. By induction on the compositions that build up $H_{0,0}^{[<h]}$, there follows the equality

$$g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(y) = \frac{P_h - P_{h-1}a_{h-1}b_h y}{Q_h - Q_{h-1}a_{h-1}b_h y}, \tag{27}$$

where P_h and Q_h are seen to satisfy the recurrence (19). Setting $y = 0$ in (27) proves (22).

Finally, $H_{h,h}^{[\geq h]}$ is determined implicitly as the root y of the equation $g_0 \circ \cdots \circ g_{h-1}(y) = H_{0,0}$, an equation that, when solved using (27), yields the form (24). $\square$

A large number of generating functions can be derived by similar techniques. See Figure 1 for a summary of what is known. We refer to the paper [**35**], where this theory was first systematically developed and to the exposition given in [**48**, Chapter 5]. Our presentation here draws upon the paper [**37**] where the theory was put to use in order to develop a formal algebraic theory of the general birth-and-death process.

We examine next a few specializations of the very general formulæ provided by Proposition 8.4 and Table 1.

EXAMPLE 11. *Standard lattice paths.* In order to count lattice paths, it suffices to effect one of the substitutions,

$$\sigma_M : a_j \mapsto z,\ b_j \mapsto z, c_j \mapsto z; \qquad \sigma_D : a_j \mapsto z,\ b_j \mapsto z, c_j \mapsto 0.$$

In the latter case level steps are disallowed, and one obtains so-called "Dyck paths"; in the former case, all three step types are taken into account, giving rise to so-called "Motzkin paths". We restrict attention below to the case of Dyck paths.

The continued fraction expressing $H_{0,0}$ is then purely periodic and hence a quadratic function:

$$H_{0,0}(z) = \cfrac{1}{1 - \cfrac{z^2}{1 - \cfrac{z^2}{1 - \ddots}}} = \frac{1}{2z^2}\left(1 - \sqrt{1-4z^2}\right),$$

since $H_{0,0}$ satisfies $y = (1 - z^2 y)^{-1}$. The families of polynomials P_h, Q_h are in this case defined by a recurrence with constant coefficients and they coincide, up to a shift of indices. Define classically the Fibonacci polynomials by the recurrence

$$F_{h+2}(z) = F_{h+1}(z) - zF_h(z), \quad F_0(z) = 0, \quad F_1(z) = 1.$$

One finds $Q_h = F_h(z^2)$ and $P_h = F_{h-1}(z^2)$. (The Fibonacci polynomials are essentially reciprocals of Chebyshev polynomials.) The GF of paths of height $\leq h$ is then given by Proposition 8.4 as

$$H_{00}^{[<h]}(z) = \frac{F_h(z^2)}{F_{h+1}(z^2)}.$$

In other words, the general results specialize to provide the algebraic part of the analysis of Catalan tree height, a topic dealt with in Chapter 7. We get more: for instance the number of ways of crossing a strip of width $h-1$ is

$$H_{0,h-1}^{[<h]}(z) = \frac{z^h}{F_h(z^2)}.$$

In the case of Dyck paths, explicit expressions result from the explicit generating functions and from the Lagrange-Bürmann inversion theorem; see Chapter 7 for details. □

EXAMPLE 12. *Area under Dyck path and coin fountains.* Consider the case of Dyck path and the parameter equal to the area below the path. Area under a lattice path can be defined as the sum of the indices (i.e., the starting altitudes) of all the variables that enter the standard encoding of the path. Thus, the BGF $D(z,q)$ of Dyck path with z marking half-length and q marking area is obtained by the substitution

$$a_j \mapsto q^j z, \quad b_j \mapsto q^j, \quad c_j \mapsto 0.$$

It proves convenient to operate with the continued fraction

$$F(z,q) = \cfrac{1}{1 - \cfrac{zq}{1 - \cfrac{zq^2}{\ddots}}}, \tag{28}$$

so that $D(z,q) = F(q^{-1}z, q^2)$. Since F and D satisfy difference equations, for instance,

$$F(z,q) = \frac{1}{1 - zqF(qz,q)}, \tag{29}$$

moments of area can be determined by differentiating and setting $q = 1$ (see Chapter 3 for a direct approach).

A general trick is effective to derive an alternative expression of F. Attempt to express the continued fraction F of (28) as a quotient $F(z,q) = A(z)/B(z)$. Then, the relation (29) implies

$$\frac{A(z)}{B(z)} = \frac{1}{1 - qz\frac{A(qz)}{B(qz)}}, \quad \text{hence } A(z) = B(qz),\ B(z) = B(qz) - qzB(q^2z),$$

where q is treated as a parameter. The difference equation satisfied by $B(z)$ is readily solved by indeterminate coefficients: this classical technique was introduced in the theory of integer partitions by Euler. With $B(z) = \sum b_n z^n$, the coefficient satisfy the recurrence

$$b_0 = 1, \quad b_n = q^n b_n - q^{2n-1} b_{n-1}.$$

This is a first order recurrence on b_n that unwinds to give

$$b_n = (-1)^n \frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)}.$$

In other words, introducing the "q-exponential function",

$$E(z,q) = \sum_{n=0}^{\infty} \frac{(-z)^n q^{n^2}}{(q)_n}, \quad \text{where} \quad (q)_n = (1-q)(1-q^2)\cdots(1-q^n), \tag{30}$$

one finds

$$F(z,q) = \frac{E(qz,q)}{E(z,q)}. \tag{31}$$

Given the importance of the functions under discussion in various branches of mathematics, we cannot resist a quick digression. The name of the q-exponential comes form the obvious property that $E(z(q-1),q)$ reduces to e^{-z} as $q \to 1^-$. The explicit form (30) constitutes in fact the "easy half" of the proof of the celebrated Rogers-Ramanujan identities, namely,

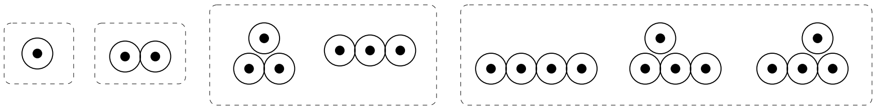
$$\begin{aligned} E(-1,q) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(q)_n} &= \prod_{n=0}^{\infty} (1-q^{5n+1})^{-1}(1-q^{5n+4})^{-1} \\ E(-q,q) = \sum_{n=0}^{\infty} \frac{q^{n(n+1)}}{(q)_n} &= \prod_{n=0}^{\infty} (1-q^{5n+2})^{-1}(1-q^{5n+3})^{-1}, \end{aligned} \tag{32}$$

that relate the q-exponential to modular forms. See Andrews' book [**4**, Ch. 7] for context.

Here is finally a cute application of these ideas to asymptotic enumeration. Odlyzko and Wilf define in [**74, 73**] an (n,m) coin fountain as an arrangement of n coins in rows in such a way that there are m coins in the bottom row, and that each coin in a higher row touches exactly two coins in the next lower row. Let $C_{n,m}$ be the number of (n,m) fountains and $C(q,z)$ be the corresponding BGF with q marking n and z marking m. Set $C(q) = C(q,1)$. The question is to determine the total number of coin fountains of area n, $[q^n]C(q)$. The series starts as (this is Sequence A005169 of the *EIS*)

$$C(q) = 1 + q + q^2 + 2q^3 + 3q^4 + 5q^5 + 9q^6 + 15q^7 + 26q^8 + \cdots,$$

as results from inspection of the first few cases.



The function $C(q)$ is *a priori* meromorphic in $|q| < 1$. From the bijection with Dyck paths and area, one finds

$$C(q) = \cfrac{1}{1 - \cfrac{q}{1 - \cfrac{q^2}{1 - \cfrac{q^3}{\ddots}}}}.$$

The identity (31) implies

$$C(q) = \frac{E(q,q)}{E(1,q)}.$$

An exponential lower bound of the form 1.6^n holds on $[q^n]C(q)$, since $(1-q)/(1-q-q^2)$ is dominated by $C(q)$ for $q > 0$. At the same time, the number $[q^n]C(q)$ is majorized by the number of compositions, which is 2^{n-1}. Thus, the radius of convergence of $C(q)$ has to lie somewhere between 0.5 and $0.61803\ldots$. It is then easy to check by numerical analysis the existence of a simple zero of the denominator, $E(-1,q)$, near $\rho \doteq 0.57614$. Routine computations based on Rouché's theorem (see Chapter 4) then makes it possible to verify formally that ρ is the only simple pole in $|q| < 3/5$, and the process is detailed in [**73**]. Thus, singularity analysis of meromorphic functions applies: *The number of coin fountains made of n coins satisfies asymptotically*

$$[q^n]C(q) = cA^n + O((5/3)^n), \quad c \doteq 0.31236, \quad A = \rho^{-1} \doteq 1.73566.$$

This example illustrates the power of modelling by continued fractions as well as the smooth articulation with meromorphic function asymptotics. □

The systematic theory of lattice path enumerations and continued fractions was developed initially because of the need to count weighted lattice paths, notably in the context of the analysis of dynamic data structures in computer science [**36**]. In this framework, a system of multiplicative weights $\alpha_j, \beta_j, \gamma_j$ is associated with the steps a_j, b_j, c_j, each weight being an integer that represents a number of "possibilities" for the corresponding step type. A system of weighted lattice paths has counting generating functions given by an easy specialization of the corresponding multivariate expressions we have just developed, namely,

(33) $$a_j \mapsto \alpha_j z, \quad b_j \mapsto \beta_j z, \quad c_j \mapsto \gamma_j z,$$

where z marks the length of paths. One can then sometimes solve an enumeration problem expressible in this way by reverse-engineering the known collection of continued fractions as found in a reference book like Wall's treatise [**99**]. Next, for general reasons, the polynomials P, Q are always elementary variants of a family of orthogonal polynomials that is determined by the weights [**21, 35, 90**]. When the multiplicities have enough structural regularity, the weighted lattice paths are likely to correspond to classical combinatorial objects and to classical families of orthogonal polynomials; see [**35, 36, 46, 48**] and Table 7 for an outline. We illustrate this by a simple example due to Lagarias, Odlyzko, and Zagier [**62**].

Objects	*Weights* $(\alpha_j, \beta_j \gamma_j)$	*Counting*	*Orth. pol.*
Simple paths	$1, 1, 0$	Catalan #	Chebyshev
Permutations	$j+1, j, 2j+1$	Factorial #	Laguerre
Alternating perm.	$j+1, j, 0$	Secant #	Meixner
Involutions	$1, j, 0$	Odd factorial #	Hermite
Set partition	$1, j, j+1$	Bell #	Poisson-Charlier
Nonoverlap. set part.	$1, 1, j+1$	Bessel #	Lommel

FIGURE 7. Some special families of combinatorial objects together with corresponding weights, moments, and orthogonal polynomials.

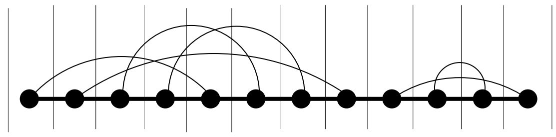


FIGURE 8. An interconnection network on $2n = 12$ points.

EXAMPLE 13. *Interconnection networks and involutions.* The problem considered here was introduced by Lagarias, Odlyzko, and Zagier in [**62**]: *There are $2n$ points on a line, with n point-to-point connections between pairs of points. What is the probable behaviour of the* width *of such an interconnection network?* Imagine the points to be $1, \dots, 2n$, the connections as circular arcs between points, and let a vertical line sweep from left to right; width is defined as the maximum number of edges encountered by such a line. One may freely imagine a tunnel of fixed capacity (this corresponds to the width) inside which wires can be placed to connect points pairwise. See Figure 8.

Let $\mathcal{I}_{2n}$ be the class of all interconnection networks on $2n$ points, which is precisely the collection of ways of grouping $2n$ elements into n pairs, or, equivalently, the class of all involutions (i.e., permutations with cycles of length 2 only). The number I_{2n} equals the "odd factorial",

$$I_{2n} = 1 \cdot 3 \cdot 5 \cdots (2n-1),$$

whose EGF is $e^{z^2/2}$ (see Chapter 2). The problem calls for determining the quantity $I_{2n}^{[h]}$ that is the number of networks corresponding to a width $\leq h$.

The relation to lattice paths is as follows. First, when sweeping a vertical line across a network, define an active arc at an abscissa as one that straddles that abscissa. Then build the sequence of active arcs counts at half-integer positions $\frac{1}{2}, \frac{3}{2}, \dots, 2n - \frac{1}{2}, 2n + \frac{1}{2}$. This constitutes a sequence of integers where each member is ± 1 the previous one, that is, a lattice path without level steps. In other words, there is an ascent in the lattice path for each element that is smaller in its cycle and a descent otherwise. One may view ascents as associated to situations where a node "opens" a new cycle, while descents correspond to "closing" a cycle.

Involutions are much more numerous than lattice paths, so that the correspondence from involutions to lattice paths is many-to-one. However, one can easily enrich lattice paths, so that the enriched objects are in one-to-one correspondence with involutions. Consider again a scanning position at a half-integer where the vertical line crosses ℓ (active) arcs. If the next node is of the closing type, there are ℓ possibilities to choose from. If the next node is of the opening type, then there is only one possibility, namely, to start a new cycle. A complete encoding of a network is obtained by recording additionally the sequence of the n possible choices corresponding to descents in the lattice path (some canonical order is fixed, for instance, oldest first). If we write these choices as superscripts, this means that the set of all enriched encodings of networks is obtained from the set of standard lattice path encodings by effecting the substitutions

$$b_j \mapsto \sum_{k=1}^{j} b_j^{(k)}.$$

The OGF of all involutions is obtained from the generic continued fraction of Proposition 8.4 by the substitution

$$a_j \mapsto z, \quad b_j \mapsto j\,z,$$

where z records the number of steps in the enriched lattice path, or equivalently, the number of nodes in the network. In other words, we have obtained combinatorially a formal continued fraction representation,

$$\sum_{n=0}^{\infty} (1 \cdot 3 \cdots (2n-1)) z^{2n} = \cfrac{1}{1 - \cfrac{1 \cdot z^2}{1 - \cfrac{2 \cdot z^2}{1 - \cfrac{3 \cdot z^2}{\ddots}}}},$$

which was originally discovered by Gauß [**99**]. Proposition 8.4 then gives immediately the OGF of involutions of width at most h as a quotient of polynomials. Define

$$I^{[h]}(z) := \sum_{n \geq 0} I_{2n}^{[h]} z^{2n}.$$

One has

$$I^{[h]}(z) = \cfrac{1}{1 - \cfrac{1 \cdot z^2}{1 - \cfrac{2 \cdot z^2}{\cfrac{\ddots}{1 - h \cdot z^2}}}} = \frac{P_h(z)}{Q_h(z)}$$

where P_h and Q_h satisfy the recurrence

$$Y_{h+1} = Y_h - hz^2 Y_{h-1}.$$

The polynomials are readily determined by their generating functions that satisfies a first-order linear differential equation reflecting the recurrence. For instance, the denominator polynomials are identified to reciprocals of the Hermite polynomials,

$$Q_h(z) = (z/2)^h H_h(\frac{1}{2z}),$$

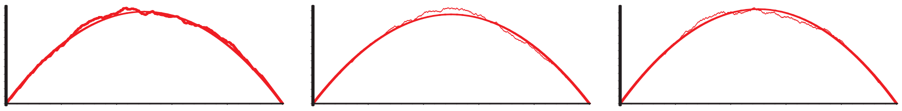


FIGURE 9. Three simulations of random networks with $2n = 1000$ illustrate the tendency of the profile to conform to a parabola with height close to $n/2 = 250$.

themselves defined classically [**2**, Ch. 22] as orthogonal with respect to the measure $e^{-x^2}\,dx$ on $(-\infty, \infty)$ and expressible as

$$H_m(x) = m! \sum_{m=0}^{\lfloor m/2 \rfloor} \frac{(-1)^j}{j!(m-2j)!}(2x)^{m-2j}.$$

In particular, one finds

$$I^{[0]} = 1, \quad I^{[1]} = \frac{1}{1-z^2}, \quad I^{[2]} = \frac{1-2z^2}{1-3z^2}, \quad I^{[3]} = \frac{1-5z^2}{1-6z^2+3z^4}, \quad \&c.$$

The interesting analysis of the dominant poles of the rational GF's, for any fixed h, is discussed in the paper [**62**]. Simulations suggest strongly that the width of a random interconnection network on $2n$ nodes is tightly concentrated around $n/2$; see Figure 9. Louchard [**67**] succeeded in proving this fact and a good deal more: With high probability, the profile (the profile is defined here as the number of active arcs as time evolves) of a random network conforms asymptotically to a deterministic parabola $x(1-x/n)$ (in the scale of n) to which are superimposed well-characterized random fluctuations of amplitude only $O(\sqrt{n})$. In particular, *the width of a random network of* $2n$ *nodes converges in probability to* $\frac{n}{2}$. □

4. Algebra of algebraic functions

Algebraic series and algebraic functions are simply defined as solutions of a polynomial equation (Definition 8.6). It is a nontrivial fact established by elimination theory (which can itself be implemented by way of resultants or Groebner bases) that they are equivalently defined as components of solutions of polynomial systems (Theorem 8.7). Like rational functions, they form a differential field but are not closed under integration; unlike rational functions, they are not closed under Hadamard products (Theorem 8.8).

The starting point is the following definition of an algebraic series.

DEFINITION 8.6. *A power series* $f \in \mathbb{C}[\![z]\!]$ *is said to be* algebraic *if there exists a (nonzero) polynomial* $P(z,y) \in \mathbb{C}[z,y]$, *such that*

$$P(z, f) = 0.$$

The set of all algebraic power series is denoted by $\mathbb{C}^{alg}[\![z]\!]$.

Rational functions correspond to the particular case where $\deg_y P(z,y) = 1$; consequently, we have $\mathbb{C}^{rat}[\![z]\!] \subset \mathbb{C}^{alg}[\![z]\!]$. The *degree* of an algebraic series f is by definition the minimal value of $\deg_y P(z,y) = 1$ over all polynomials that are cancelled by f (so that rational series are algebraic of degree 1). Note that one can always assume P to be irreducible (that is $P = QR$ implies that one of Q or R is a scalar) and of minimal degree. In effect, assume that $P(z, f(z)) = 0$ where P factorizes: $P(z,y) = P_1(z,y)P_2(z,y)$. We must have at least one of the equalities $P_1(z,f) = 0$ or $P_2(z,f) = 0$, since otherwise, one would have two nonzero formal power series $g_1 = P_1(z, f)$, $g_2 = P_2(z, f)$ such that $g_1 \cdot g_2 = 0$, which is absurd.

We do not address at this stage the question of deciding whether a bivariate polynomial defines one, or several algebraic series, or none. The question is completely settled by means of the Newton polygon algorithm discussed in the next section.

EXERCISE 25. (i) Prove that there is, up to constant multiples, a unique polynomial of minimal degree that is cancelled by an algebraic series $f \in \mathbb{C}^{\text{alg}}[\![z]\!]$.

(ii) Prove furthermore that if f lies in $\mathbb{Q}[\![z]\!]$, then the minimal polynomial may be chosen in $\mathbb{Q}[z, y]$. (Thus, for enumerative problems, factorization over the field $\mathbb{Q}$ is all that is ever required; there is no need for "absolute" factorization, that is, factorization over the complex numbers.)

[Hint. Point (i) is related to gcd's and principal ideal domains. Point (ii) is a classical lemma of Eisenstein; see [**13**].]

4.1. Characterization and elimination. A polynomial system is by definition a set of equations

$$\left\{ \begin{array}{lcl} P_1(z, y_1, \ldots, y_m) & = & 0 \\ \vdots & & \\ P_m(z, y_1, \ldots, y_m) & = & 0, \end{array} \right. \tag{34}$$

where each P_j is a polynomial. A solution over $\mathbb{C}[\![z]\!]$ of (34) is an m-tuple $(f_1, \ldots, f_m) \in \mathbb{C}[\![z]\!]^m$ that cancels each P_j, that is, $P_j(z, f_1, \ldots, f_m) = 0$. Any of the f_j is called a component solution. A basic result is that any component solution of a (nontrivial) polynomial system is an algebraic series. In other words, one can eliminate the auxiliary variables $y_2, \ldots, y_m$ and construct a single bivariate polynomial Q such that $Q(z, y_1) = 0$. This result is a famous one in the theory of polynomial systems.

The study of polynomial systems and algebraic varieties (i.e., the point set of all solutions of a polynomial system) involves some subtle mathematics. Although generic instances (including well-posed combinatorial problems) are normally well-behaved, all sorts of degeneracies might occur in principle, and these have to be accounted for or disposed of in statements of theorems. For instance, the number π is known to be a transcendental number, i.e., it does not satisfies any polynomial equation with coefficients in $\mathbb{Q}$. However, the triple $(1, 1, \pi)$ is a solution of the following polynomial system in (y_1, y_2, y_3):

$$y_1 - 1 = 0,\ y_2 + y_1^2 - 2 = 0,\ (y_1 - y_2)y_3 = 0. \tag{35}$$

The intuitive reason is that the system is pathological: it does not specify a *finite* number of values since the equation binding y_3 is algebraically "equivalent" to $0y_3 = 0$.

As this example (35) suggests, in order to avoid degeneracies, we shall restrict attention to systems of polynomials that have a finite set of solutions (in the algebraic closure of the coefficient field). Technically, such systems are called *zero-dimensional*[5]. It is a nonobvious fact that the property for a system to be zero-dimensional is algorithmically checkable (by means of Groebner bases). In the statement that follows, zero-dimensionality is meant for the field of coefficients $\mathbb{C}(z)$ and for algebraic varieties defined in the algebraic closure of $\mathbb{C}(\!(z)\!)$.

[5]For the general notion of dimensionality based on Hilbert polynomials, consult for instance [**27**, §9.3].

THEOREM 8.7 (Algebraic function characterization and elimination). *For a power series $f(z) = \sum_n f_n z^n$ in $C[\![z]\!]$, the following conditions are equivalent to algebraicity:*

(i) *Normal form: there exists an irreducible polynomial $P(z, y) \in \mathbb{C}[z, y]$ such that*

$$P(z, f(z)) = 0.$$

(ii) *Elimination: there exists a zero-dimensional system of polynomials $\{P_j(z, \mathbf{y})\}_{j=1}^m$ such that a solution $\mathbf{y}(z) \in \mathbb{C}[\![z]\!]^m$ to the system*

$$\{P_j(z, y_1, y_2, \ldots, y_m) = 0\}, \qquad j = 1 \,..\, m, \tag{36}$$

has $y_1 = f$.

Proof. That $(i) \implies (ii)$ is immediate, since a single equation is a particular system of dimension $m = 1$ and a single polynomial of degree d has at most d roots in any extension field so that the variety it defines is 0-dimensional. The converse property $(ii) \implies (i)$ expresses the fact that auxiliary variables can always be eliminated from a system of polynomial equations. We sketch below an abstract algebraic argument excerpted from [**27**].

A basic principle in the theory of equations consists in examining the collection of all the "consequences" of a given system of equations. This is formalized by the notion of *ideal*. Given a field $\mathbb{K}$, an ideal of the polynomial ring $R = \mathbb{K}[y_1, \ldots, y_m]$ is a set I that is closed under sums ($a, b \in I$ implies $a + b \in I$) and under multiplication by arbitrary elements of R ($a \in I$ and $c \in R$ imply $ca \in I$). The ideal $\langle h_1, \ldots, h_s \rangle$ generated by the $h_j \in R$ is by definition the set of all combinations $\sum c_j h_j$ for arbitrary $c_j \in R$.

Consider a system of s polynomial equations

$$(\Sigma) \qquad \{h_j(y_1, \ldots, y_m) = 0\}_{j=1}^s .$$

Set $\vec{y} = (y_1, \ldots, y_m)$. Clearly if it happens that some particular value of $\vec{y}$ satisfies the system Σ,, then this value also cancels any polynomial g in the ideal $\langle h_1, \ldots, h_s \rangle$. Now the intersection

$$I_1 = \mathbb{K}[y_1] \cap \langle h_1, \ldots, h_s \rangle,$$

is an ideal (by elementary algebra, the intersection of two ideals is an ideal). The ideal I_1 consists of a collection of univariate polynomials which vanish at any y_1 that is a component of a solution of Σ. In $\mathbb{K}[y_1]$, all ideals are principal, meaning that for some polynomial g, one has $I_1 = \langle g \rangle$. (The last fact relies simply on the existence of a Euclidean division, hence of gcd's in $\mathbb{K}[y_1]$.) Thus, there exists a polynomial g such that $g(y_1) = 0$. (In passing, one needs to argue that g is not the null polynomial. However, if this was the case, it would contradict the assumption of zero-dimensionality.) Finally, specializing the discussion to the coefficient field $\mathbb{K} = \mathbb{C}(z)$, i.e., treating z as a parameter, then yields the elimination statement. □

Note that by the famous Hilbert basis theorem [**27**, p. 74], every polynomial ideal has a finite generating set ("basis"). As a consequence, the intersection construction used in the proof of Theorem 8.7 can be extended to obtain a triangular system $\{g_j = 0\}$, where each g_j depends on $y_1, \ldots, y_j$ alone. This makes it possible to obtain eventually all components of all solution vectors by successively solving for y_1, then y_2, etc.

Techniques to perform eliminations can be made "effective" (i.e., algorithmic), two major methods being resultants and Groebner bases. A complete exposé of elimination theory is however beyond the scope of the book. We shall thus limit ourselves to sketching a strategy based on resultants that is conceptually simple. We shall next illustrate informally the fundamental principles of Groebner bases by way of example. For a comprehensive discussion of these issues, we refer to the excellent introduction given by Cox, Little, and O'Shea [**27**].

Resultants. Consider a field of coefficients $\mathbb{K}$ which may be specialized as $\mathbb{Q}, \mathbb{C}, \mathbb{C}(z), \mathbb{C}((z))$, as the need arises. A polynomial of degree d in $\mathbb{K}[x]$ has at most d roots in $\mathbb{K}$ and exactly d roots in the algebraic closure $\bar{\mathbb{K}}$ of $\mathbb{K}$. First, we set:

DEFINITION 8.7. *Given two polynomials,*

$$P(x) = \sum_{j=0}^{\ell} a_j x^{\ell-j}, \qquad Q(x) = \sum_{k=0}^{m} b_{m-k} x^k,$$

their resultant *(with respect to the variable x) is the determinant of order $(\ell + m)$:*

$$\mathbf{R}(P, Q, x) = \det \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & 0 & 0 \\ 0 & a_0 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{\ell-1} & a_\ell \\ b_0 & b_1 & b_2 & \cdots & 0 & 0 \\ 0 & b_0 & b_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{m-1} & b_m \end{vmatrix}. \tag{37}$$

The matrix itself is often called the Sylvester matrix and the resultant is also known as the Sylvester determinant. By its definition, the resultant is a polynomial form in the coefficients of P and Q. The main property of resultants is the following.

LEMMA 8.2. (i) *If $P(x), Q(x) \in \mathbb{K}[x]$ have a common root in the algebraic closure $\bar{\mathbb{K}}$ of $\mathbb{K}$, then*

$$\mathbf{R}(P(x), Q(x), x) = 0.$$

(ii) *Conversely, if $\mathbf{R}(P(x), Q(x), x) = 0$ holds, then either $a_0 = 0$ or $b_0 = 0$ or else $P(x), Q(x)$ have a common root in $\bar{\mathbb{K}}$.*

Proof. We refer globally to [**64**, V§10] for a presentation of resultants.

(i) If P and Q have 0 as a common root, then by construction, the resultant is equal to 0 since the determinant has its last column equal to the 0-vector. Let S be the Sylvester matrix of P, Q. If P and Q have a common root ξ with $\xi \neq 0$ then, the linear system

$$S \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{\ell+m-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{38}$$

admits a non trivial solution, $(w_0, \ldots, w_{\ell+m-1})$ with $w_j = \xi^{l+m-1-j}$. This can be seen since the Sylvester matrix is by construction the matrix of coefficients of the collection of polynomials

$$x^{m-1}P(x), \ldots, xP(x), P(x);\ x^{\ell-1}Q(x), \ldots, xQ(x), Q(x).$$

(In fact, the argument above provides the rationale for the definition (37).) Now, by Cramer's rule[6], the existence of a nontrivial solution to the linear system (38) implies that the Sylvester determinant has to be 0.

(ii) Conversely, let $\alpha_1, \ldots, \alpha_\ell$ and $\beta_1, \ldots, \beta_m$ be respectively the roots of P and Q in the algebraic closure $\bar{\mathbb{K}}$. It is a known fact [**64**] that the resultant $\mathcal{R}$ of P, Q satisfies

$$\mathcal{R} = a_0^\ell b_0^m \prod_{i,j}(\alpha_i - \beta_j) = a_0^m \prod_i Q(\alpha_i) = (-1)^{\ell m} b_0^\ell \prod_j P(\beta_j). \tag{39}$$

Thus, the fact that P, Q have a common root implies $R = 0$. □

EXERCISE 26. Let $f(x)$ be a polynomial of degree ℓ with leading coefficient a_0 and m distinct roots $\alpha_1, \ldots, \alpha_\ell$. Then

$$\mathbf{R}(f(x), f'(x), x) = a_0^\ell \prod_{i \neq j}(\alpha_i - \alpha_j).$$

Let $g(x)$ have possibly multiple roots. Develop an algorithm based on gcd computations and resultants to produce a lower bound on the distance between any two distinct roots of g (a "separation distance").

The resultant thus provides a *sufficient* condition for the existence of common roots, but not always a necessary one. This has implications in situations where the coefficients a_j, b_k depend on one or several parameters. In that case, the condition $\mathbf{R}(P, Q, x) = 0$ will certainly capture all the situations where P and Q have a common root, but it may also include some situations where there is a reduction in degree, although the polynomials have no common root. For instance, taking $P = tx - 1, Q = tx^2 - 1$ (with t a parameter), the resultant with respect to x is found to be

$$\mathcal{R} = t(1 - t).$$

Indeed, the condition $\mathcal{R} = 0$ corresponds to either a common root ($t = 1$ implying $P(1) = Q(1) = 0$) or to some degeneracy in degree ($t = 0$). (Note for this discussion that the resultant applies in particular when the base field is $\mathbb{Q}(t)$ or $\mathbb{C}(t)$ and the algebraically closed field contains $\mathbb{Q}((t))$ or $\mathbb{C}((t))$, which is the case of interest what follows.)

Given a system (36), we can then proceed as follows in order to extract a single equation satisfied by one of the indeterminates. By taking resultants with P_m, eliminate all occurrences of the variable y_m from the first $m - 1$ equations, thereby obtaining a new system of $m - 1$ equations in $m - 1$ variables (and z kept as a parameter, so that the base field is $\mathbb{C}(z)$). Repeat the process and successively eliminate $y_{m-1}, \ldots, y_2$. The strategy (in the simpler case where variables are eliminated in succession exactly one at a time) is summarized in the procedure eliminateR of Figure 10.

The algorithm eliminateR of Figure 10 will, by virtue of the sufficiency of resultant conditions, always provide a correct polynomial condition satisfied by y_1, although the resulting polynomial that cancels y_1 need not be minimal. In short, elimination by resultants produces valid but not necessarily minimal results.

Alternatively, one can appeal to the theory of Groebner bases in order to develop a complete algorithm. The Groebner based algorithm is summarized as the procedure eliminateG in Figure 10 and we refer to the already cited book [**27**] for precise definitions and complete detail. See also Example 16 below for a presentation by way of example.

[6]This reasoning is valid in any field of characteristic 0 since only rational operations are used in the proof.

procedure eliminateR $(P_1, \ldots, P_m, y_1, y_2, \ldots y_m)$;
$\{$*Elimination of* $y_2, \ldots, y_m$ *by resultants*$\}$
$(A_1, \ldots, A_m) := (P_1, \ldots, P_m)$;
for j **from** m **by** -1 to 2 **do**
for k **from** $j-1$ **by** -1 to 1 **do**
$A_k := \mathbf{R}(A_k, A_j, y_{k+1})$;
return(A_1).

procedure eliminateG $(P_1, \ldots, P_m; y_1, y_2, \ldots y_m)$;
$\{$*Elimination of* $y_2, \ldots, y_m$ *by Groebner bases*$\}$
Fix the *variable ordering* $y_m > y_{m-1} > y_2 > y_1$;
Choose the *pure lexicographic ordering* on monomials;
Operate with the coefficient field $\mathbb{C}(z)$;
Construct a *Groebner basis* G of the ideal $\langle P_1, \ldots, P_m\rangle$.
return$(G \cap \mathbb{C}(z)[y_1])$

FIGURE 10. Resultant elimination (top) and Groebner basis elimination (bottom).

Computer algebra systems usually provide implementations of both resultants and Groebner bases. The complexity of elimination is however exponential in the worst-case. Degrees essentially multiply; this is somewhat intrinsic since y_0 in the system

$$y_0 - z - y_k = 0,\ y_k - y_{k-1}^2 = 0,\ \ldots\ , y_1 - y_0^2 = 0,$$

defines with k equations the GF of regular trees of degree 2^k, while it represents an algebraic function of degree 2^k and no less.

The example of coloured trees. We illustrate the previous discussion by means of a combinatorial problem that is treated in succession by a "pencil-and-paper" approach (that makes use of a simple combinatorial property and high-school algebra), then by resultants and finally by Groebner bases of which the last case serves to demonstrate the *modus operandi* of Groebner bases.

EXAMPLE 14. *Three-coloured trees—pencil-and-paper approach.* Consider (plane, rooted) binary trees with nodes coloured by any of three colours, a, b, c. We impose that the external nodes are coloured by the a-colour. The colouring has to be perfect in the sense that no two adjacent nodes are assigned the same colour. We let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ denote the set of perfectly coloured trees with root of the a, b, c type respectively. Let A, B, C be the corresponding OGF's where size is taken to be the number of external nodes. The defining equations are thus:

$$\text{(40)}\qquad \left\{\begin{array}{lcl} \mathcal{A} & = & a + (\mathcal{B}+\mathcal{C})^2 \\ \mathcal{B} & = & (\mathcal{C}+\mathcal{A})^2 \\ \mathcal{C} & = & (\mathcal{A}+\mathcal{B})^2 \end{array}\right. \qquad \left\{\begin{array}{lcl} A - z + (B+C)^2 & = & 0 \\ B - (C+A)^2 & = & 0 \\ C - (A+B)^2 & = & 0. \end{array}\right.$$

The pencil-and-paper approach may start by observing that $\mathcal{B}$ and $\mathcal{C}$ are isomorphic as combinatorial classes, as a result of the bijection that exchanges the b and c colours. Thus

$B = C$. Then equation (40) simplifies into a system of 2 equations in 2 unknowns with z a parameter:

$$\left\{\begin{array}{lcl} A - z - 4B^2 & = & 0 \\ B - B^2 - 2AB - A^2 & = & 0. \end{array}\right. \tag{41}$$

Since B is to be eliminated, we may combine linearly the equations of (41) in order to get rid of B^2 (the highest power in B),

$$\left\{\begin{array}{lcl} A - z - 4B^2 & = & 0 \\ A - z - 4B + 4A^2 + 8AB & = & 0 \end{array}\right. \tag{42}$$

The second equation is linear in B and the value can then be plugged back into the first equation, resulting in a fourth degree polynomial that cancels A,

$$R(z, A) = 16\,A^4 - 8\,A^3 + (17 + 8\,z)\,A^2 - (4 + 18\,z)\,A + z^2 + 4\,z\,. \tag{43}$$

and the polynomial is easily checked to be irreducible. ☐

EXAMPLE 15. *Three-coloured trees—resultant approach.* Start again with the system (40) and take mechanically resultants with respect to C of the third equation together with the first and second equations. This eliminates C and provides two polynomials cancelled by A and B:

$$\begin{array}{l} B - A^2 - 2\,A^3 - A^4 - 4\,A^3B - 6\,A^2B^2 - 4\,A^2B - 4\,AB^3 - 2\,B^2A - B^4 \\ A - z - B^2 - 2\,A^2B - A^4 - 4\,A^3B - 6\,A^2B^2 - 4\,B^2A - 4\,AB^3 - 2\,B^3 - B^4. \end{array} \tag{44}$$

A further resultant with respect to B will then yield a polynomial involving A alone and cancelled by A:

$$\begin{array}{l} 256\,A^8 + 256\,A^7 + (256\,z + 352)\,A^6 - (64\,z - 304)\,A^5 \\ + \left(96\,z^2 - 112\,z + 161\right)A^4 - \left(80\,z^2 + 108\,z - 26\right)A^3 + \left(16\,z^3 - 26\,z^2 - 22\,z - 7\right)A^2 \\ - \left(12\,z^3 + 10\,z^2 + 2\,z + 4\right)A + 9\,z^2 + 6\,z^3 + 4\,z + z^4 \end{array}$$

This time, the polynomial of degree 8 in A. (From (43), we know it cannot be minimal.) In fact it factors as

$$(4A^2 + 3A + 1 + z)^2\,(16A^4 + \cdots + 4z) = (4A^2 + 3A + 1 + z)^2\,R(z, A), \tag{45}$$

with $R(z, A)$ as in (43).

The series expansion for the combinatorial generating function A starts with

$$A(z) = z + 4z^4 + 16z^5 + 56z^6 + 256z^7 + 1236z^8 + 5808z^9 + \cdots\,,$$

so that the first quadratic factor is disqualified as a candidate for minimal polynomial. We have thus found again the result of (43). Note that the complete factorization of (45) over $\mathbb{Q}$ guarantees at the same time that $R(z, A)$ is the minimal polynomial so that A is of exact degree 4. (See the exercise preceding Theorem 8.7.) ☐

EXAMPLE 16. *Three-coloured trees—Groebner basis approach.* In accordance with the procedure eliminateG we first fix an ordering on variables. For the case at hand, we choose

$$C > B > A,$$

with the intent that variables should be eliminated in that order, first C, next B, then A. Several orderings may be defined on monomials, but for our purposes, we chose the *pure* *lexicographic ordering* that is the usual alphabetic ordering of words in a dictionary (with C preceding B preceding A, though!). For instance, with this ordering, we have (think of $C = u$, $B = v$, $A = w$ and the lexicographic order in the Latin script)

$$C > C^2 > C^3 > CB > CB^2 > B > B^5 > BA > BA^4.$$

The leading term of a polynomial with the chosen ordering is denoted by $\operatorname{LT}(p)$. For instance: $\operatorname{LT}(2A^2 + 3B^2 + 4BA + 5CB^2) = 5CB^2$.

The basic operation of Groebner basis theory is the "division" operation of a polynomial f by a sequence of polynomials $\mathcal{F} = (f_1, \ldots, f_s)$ producing a remainder denoted by $\overline{f}^{\mathcal{F}}$. The idea is to "reduce" the polynomial f by an operation that is reminiscent of Gaussian elimination (for multivariate linear polynomials) and of the Euclidean algorithm (for univariate nonlinear polynomials).

The principle of division is as follows. First transform each polynomial in f_i of $\mathcal{F}$ into a rewrite rule:

$$(W_i): \qquad \operatorname{LT}(f_i) \mapsto -f_i + \operatorname{LT}(f_i).$$

(a rule thus replaces a monomial by a combination of smaller monomials under the monomial order.) Use repeatedly $W_1, \ldots, W_s$ (in that order, say, for determinacy) to reduce the leading term of f (that changes at each stage). When the leading term no longer reduces, then proceed similarly with the second leading term, the third one, and so on. The algorithm is specified in full in [**27**, p. 62]. On our example, the three rules are

$$\begin{array}{llll} (W_1) & C^2 & \mapsto & -z + A - B^2 - 2BC \\ (W_2) & C^2 & \mapsto & 2CA + A^2 \\ (W_3) & C & \mapsto & A^2 + 2AB + B^2 \end{array}$$

An instance of a reduction will be:

$$\begin{aligned} C^3 &\overset{(W_1)}{\mapsto} CA - CB^2 - 2BC^2 - zC \\ &\overset{(W_1)}{\mapsto} CA + 3CB^2 - Cz - 2AB + 2B^3 + 2Bz \overset{(W_3)}{\mapsto} \cdots . \end{aligned}$$

(In such a simple case, the effect of the rules is easy to describe: they eventually clear out all occurrences of C, with W_1 used a number of times and W_3 used once at the end to dispose of the last linear term in C.)

Next comes the notion of *Groebner basis*. The principle is to build a "good" set of generators for the *ideal* defined by $\mathcal{F}$, that is, the set of all polynomials

$$I = \sum_j a_j f_j,$$

where the coefficients a_j are polynomials. (On the example, we would have $a_j \in \mathbb{C}(z)[A, B, C]$.)

DEFINITION 8.8. *A Groebner basis for an ideal I is a set $\mathcal{G}$ of polynomials such that for all $f \in I$, the division by $\mathcal{G}$ yields a complete reduction,*

$$\overline{f}^{\mathcal{G}} = 0.$$

In particular, such a basis should "self-reduce" in the sense that for all $g, h \in \mathcal{G}$:

$$\overline{x^\alpha g - x^\beta h}^{\mathcal{G}} = 0. \tag{46}$$

(Here x^α and x^β represent arbitrary monomials, in accordance with standard multi-index notation.)

Buchbergers' algorithm constructs Groebner bases by enriching a given base $\mathcal{F}$ till the self-reduction property (46) is ensured. Define the S-polynomial of two monic polynomials p, q as $x^\alpha p - x^\beta q$, with α and β chosen to be the smallest monomials that achieve cancellation of the leading terms of p and q. (S-polynomials are in a sense the simplest nontrivial polynomials that should reduce.) One must have $\overline{S(g,h)}^{\mathcal{G}} = 0$, for all $g, h \in \mathcal{G}$. Then, the construction of a Groebner basis from a sequence $\mathcal{F}$ proceeds incrementally: we start with $\mathcal{H} = \mathcal{F}$ and successively add to $\mathcal{H}$ all pairs p, q with $p, q \in \mathcal{H}$ such that $\overline{S(p,q)}^{\mathcal{H}} \neq 0$. The process is stopped once saturation is reached. (The algorithm is specified in full in [**27**, p. 97].)

For instance consider the system (40), that is, $\mathcal{F} = (f_1, f_2, f_3)$. Set initially $\mathcal{H} = \mathcal{F}$. The S-polynomial of f_1, f_2 is obtained by taking $1 f_1 - 1 f_2$ (the monomials C^2 disappear), giving

$$-A + B^2 + 2BC + z + B - 2CA - A^2,$$

which reduces modulo $\mathcal{H}$ (by way of (W_3)) into

$$f_4 = B - A^2 - A + B^2 + z - 2A^2B + 2AB^2 + 2B^3 - A^3.$$

Then, the two other pairs provide $f_5 = \overline{f_2 f_3}^{\mathcal{H}} \neq 0$, and $f_6 = \overline{f_3 f_1}^{\mathcal{H}} \neq 0$. At this stage, the new-comers f_4, f_5, f_6 are adjoined to $\mathcal{H}$, giving the new value $\mathcal{H} = (f_1, \ldots, f_6)$. The completion process continues till $\mathcal{H}$ eventually stabilizes. We then take the Groebner basis to be $\mathcal{G} := \mathcal{H}$, after removing redundant polynomials [**27**, Sec. 2.9]. Here, a set of 4 polynomials is obtained. In accordance with theory, the Groebner basis contains a single polynomial that depends on the variable A alone; this polynomial that is of degree 6 factors as

$$(4A^2 + 3A + z + 1)R(z, A).$$

The minimal polynomial is once more recovered: Groebner elimination has faithfully accomplished its task. □

4.2. Closure properties and coefficients. By definition, algebraic functions satisfy strong algebraic closure properties. The proofs now fall as ripe fruits as a benefit of our investment on elimination. From this point on, eliminate will designate either of the eliminateR or eliminateG procedure.

THEOREM 8.8 (Algebraic function closure (1)). *The set $\mathbb{C}^{alg}[[z]]$ of algebraic series is closed under the operations of sum $(f + g)$, product $(f \times g)$, quasi-inverse (defined by $f \mapsto (1 - f)^{-1}$, conditioned upon $f_0 = 0$), differentiation (∂_z), composition $(f \circ g$, conditioned upon $g_0 = 0)$.*

Proof. Let v and w be defined by $P(z, u) = 0$ and $Q(z, v) = 0$. For sum, product, quasi-inverse, and composition, just appeal to

$$\begin{array}{ll} y = u + v : & \mathsf{eliminate}([y - u - v, P(z,u), Q(z,v)],\ y, u, v); \\ y = uv : & \mathsf{eliminate}([y - uv, P(z,u), Q(z,v)],\ y, u, v); \\ y = (1-u)^{-1} : & \mathsf{eliminate}([y - uy - 1, P(z,u)],\ y, u); \\ y = u \circ v(z) : & \mathsf{eliminate}([P(v,y), Q(z,v)],\ y, v). \end{array}$$

An immediate consequence is that if $U(z, y)$ is a rational function and $\alpha \in \mathbb{C}^{\text{alg}}[[z]]$ is algebraic, then $U(z, \alpha)$ is algebraic. Next, if α satisfies $P(z, \alpha) = 0$, then the derivative

α' with respect to z satisfies

$$P'_z(z,\alpha) + \alpha' P'_y(z,\alpha) = 0 \quad \text{or} \quad \alpha' = -\frac{P'_z(z,\alpha)}{P'_y(z,\alpha)},$$

and is therefore a rational fraction in α. Thus, by the previous observation, α' is algebraic. $\square$

These results show that $\mathbb{C}^{\text{alg}}[[z]]$ is a ring. They extend transparently to $\mathbb{C}^{\text{alg}}((z))$ which is a field.

Algebraic series are not closed under Hadamard product. To see it, consider

$$h(z) := \frac{1}{\sqrt{1-z}} \odot \frac{1}{\sqrt{1-z}} = \sum_{n=0}^{\infty} \frac{1}{16^n} \binom{2n}{n}^2 z^n.$$

The coefficient $[z^n]h(z)$ is asymptotic to $1/(\pi n)$. Consequently, by an Abelian theorem,

$$h(z) \sim \frac{1}{\pi} \log \frac{1}{1-z}, \qquad z \to 1^-.$$

Such a singularity is incompatible with algebraicity, where only fractional power series may occur; see the next section. On the positive side, we have the following.

THEOREM 8.9 (Algebraic function closure (2)). (i) *If f is rational and g is algebraic, then the Hadamard product $f \odot g$ is algebraic.*

(ii) *Define the diagonal of a bivariate series $F \in \mathbb{C}[x,y]$,*

$$\Delta_{x,y} F(x,y) = \sum_n F_{n,n} z^n;$$

Then, the diagonal of a rational bivariate series is an algebraic series.

Proof. (i) Take $f \in \mathbb{C}^{\text{rat}}[[z]]$ and $g \in \mathbb{C}^{\text{alg}}[[z]]$ and consider the Hadamard product $h = f \odot g$. The Hadamard product distributes over sums, so that is is enough to show algebraicity when f is a simple fraction element, $f = (1-az)^{-m}$. This is settled by simple algebra, since

$$(1-az)^{-m} \odot g(z) = (1-z)^{-m} \odot g(az) = \frac{d^{m-1}}{dz^{m-1}} \left(z^{m-1} g(az)\right).$$

An alternative, more analytic, derivation can be based on Hadamard's formula for Hadamard products. This important formula valid for all functions analytic at 0 is

$$f \odot g(z) = \frac{1}{2i\pi} \int_\gamma f(t) g(\frac{z}{t}) \frac{dt}{t}. \tag{47}$$

If f and g are analytic in $|z| < R$ and $|z| < S$ respectively, then γ can be any loop around zero whose elements t satisfy $|t| < R$ and $|z/t| < S$, that is,

$$|z| S < |t| < R,$$

and the conditions are compatible provided z is of small enough modulus. (The verification by direct series expansions is immediate.) For the original problem of $\mathbb{C}^{\text{rat}}[[z]] \odot \mathbb{C}^{\text{alg}}[[z]]$, an evaluation of (47) by residues then provides the result.

(ii) Our proof will start from another of Hadamard's formulæ,

$$\Delta_{x,y} F(x,y) = \frac{1}{2i\pi} \int_\gamma F(t, \frac{z}{t}) \frac{dt}{t}. \tag{48}$$

that is proved by the same devices as (47). If $F(x,y)$ is analytic in $|x| < R$ and $|y| < S$, then t should be taken such that $|z| S < |t| < R$, and the domain of possible values is nonempty as long as $|z|$ is small enough. (The verification is again by straight expansion.) This formula generalizes (47) since $f \odot g(z) = \Delta_{x,y}(f(x)g(y))$.

Now, for $|z|$ small, we can evaluate the integral by residues inside $z/t < S$. See [**88**] for details. Assume z small enough and let $F(x,y) = A(x,y)/B(x,y)$. Consider the roots $\alpha_1(z), \ldots, \alpha_r(z)$ of the characteristic equation, $B(\alpha, z/\alpha) = 0$ that tend to 0 as $z \to 0$. Then, for z small enough and γ encircling the "small" roots $\alpha_1(z), \ldots, \alpha_r(z)$, a residue computation gives

$$\Delta_{x,y} = \sum_{j=1}^{r} \operatorname{Res}\left(F(t, \frac{z}{t})\right)_{t=\alpha_j(z)}.$$

Thus, the diagonal is a rational function of (some) branches of an algebraic function and it is therefore rational. $\square$

Closed form for coefficients. Coefficients of algebraic functions satisfy interesting relations, starting with recurrences of finite order. In addition, they can be presented as combinatorial sums.

First, we show that algebraic functions satisfy differential equations. Take $\alpha \in \mathbb{C}^{\text{alg}}[\![z]\!]$ an algebraic series of degree d, with minimal polynomial $P(z,y)$; and let $U(z,y)$ be a bivariate rational function. We have seen before that $U(z,\alpha)$ is also algebraic. In fact, we are going to show that U can be rewritten as

$$U(z,\alpha) = \sum_{j=0}^{d-1} q_j(z)\alpha^j, \qquad q_j(z) \in \mathbb{C}(z).$$

In other words, a quantity $C(z,\alpha)$ with α algebraic is representable by a polynomial of $\mathbb{C}(z)[\alpha]$. The situation generalizes what is known about algebraic numbers and reduction to polynomial form, for instance,

$$\frac{1}{\sqrt[3]{2}+\sqrt{3}} = \frac{1}{3}\sqrt[3]{2^2} - \frac{1}{6}\sqrt[3]{2^2}\sqrt{3} + \frac{1}{6}\sqrt[3]{2} - \frac{1}{6}\sqrt[3]{2}\sqrt{3} + \frac{1}{3}.$$

Write $U(z,y) = V(z,y)/W(z,y)$. Clearly, it suffices to prove that $1/W(z,\alpha)$ reduces to polynomial form. We may freely assume that W is not a multiple of P, and since P is irreducible, we have $\gcd(P,W) = 1$, where the gcd is taken in $\mathbb{C}(z)[y]$. Then, by Bézout's theorem, there exists $a(z,y), b(z,y) \in \mathbb{C}(z)[y]$ such that $aW - bP = 1$. In other words, a is the inverse of W modulo P. We thus have, at the expense of a single gcd and simple polynomial reduction:

$$U(z,\alpha) \equiv [a(z,y)V(z,y) \bmod P(z,y)]_{y=\alpha}.$$

This gives the following basic result: *Every rational fraction in z and α lives in the set $\mathbb{C}[z](\alpha)_{<d}$ of polynomials with α-degree bounded from above by d.*

Consider now $C(z)[y]_{<d}$ as a $C(z)$–vector space of dimension d. The $(d+1)$ first derivatives

$$\alpha(z),\ \alpha'(z),\ \ldots,\ \alpha^{(d)}(z)$$

lie in that space and thus they must be bound. As a consequence, there exist coefficients c_j (in $\mathbb{C}(z)$) such that

$$\sum_{j=0}^{d} c_j(z)\frac{d^j}{dz^j}\alpha(z) = 0.$$

This relation can be obtained constructively by cancelling a minor of maximal rank of the determinant expressing the linear dependency[7] Extracting coefficients then produces a recurrence of the form

$$\sum_{j=0}^{e} d_j(n)\alpha_{n+j} = 0,$$

for a sequence of coefficients $d_j(n) \in \mathbb{C}(n)$. (Denominators can always be eliminated so that the coefficients can be taken to be polynomials of $C[n]$.)

This discussion brings us to a general statement about coefficients of algebraic functions.

THEOREM 8.10 (Algebraic coefficients). (i) *The coefficients of any algebraic power series $y(z)$ satisfy a linear recurrence with polynomial coefficients, namely,*

$$\sum_{j=0}^{e} d_j(n)y_{n+j} = 0,$$

where $y_n = [z^n]y(z)$ and $c_j(n) \in \mathbb{C}[n]$.

(ii)*Let $\Phi(z,y)$ be a bivariate polynomial such that $\Phi(0,0) = 0$, $\Phi'_y(0,0) = 0$ and $\Phi(z,0) \neq 0$. Consider the algebraic function implicitly defined by $f(z) = \Phi(z, f(z))$. Then, the Taylor coefficients of $f(z)$ admit expressions as combinatorial sums,*

$$[z^n]\, f(z) = \sum_{m\geq 1} \frac{1}{m}[z^n y^{m-1}]\Phi^m(z,y). \tag{49}$$

Proof. Part (i) is established by the discussion above. Part (ii) is based on a yet unpublished memo of Flajolet and Soria, from which the discussion that follows is extracted.

First, a variant form of the coefficients is available, and Eq. (49) is elementarily equivalent to

$$[z^n]\, f(z) = \sum_{m} [z^n y^{m-1}]\Big(\Phi^m(z,y)\big(1 - \Phi'_y(z,y)\big)\Big), \tag{50}$$

as results from the observation that $(m+1)g'(y)g^m(y) = \frac{d}{dy}g^{m+1}(y)$.

The starting point of the analytic part of the proof is an integral formula giving the root of an equation $\phi(y) = 0$ inside a simple domain. Let $\phi(y)$ be analytic and assume that inside the domain defined by a closed curve γ, the equation $\phi(y) = 0$ has a unique root. Then, we have

$$\text{RootOf}(\phi(y) = 0, \gamma) = \frac{1}{\pi}\int_\gamma y\frac{\phi'(y)}{\phi(y)}\, dy. \tag{51}$$

That formula can be seen as a modified form of the "principle of the argument" ([**51**]), but it can also be checked directly via a residue computation.

The algebraic function $f(z)$ under consideration is a root of $y - \Phi(z,y) = 0$. Therefore, if we proceed formally and apply formula (51) to $\phi(y) \equiv y - \Phi(z,y)$, we get

$$\begin{aligned} f(z) &= \frac{1}{2i\pi}\int_\gamma y\frac{1-\Phi'_y(z,y)}{y-\Phi(z,y)}\, dy \\ &= \frac{1}{\pi}\int_\gamma \frac{1-\Phi'_y(z,y)}{1-\frac{1}{y}\Phi(z,y)}\, dy. \end{aligned} \tag{52}$$

[7]This result is often referred to in the combinatorics literature as "Comtet's theorem"; see [**25**].

Still proceeding formally, we expand the integrand using

$$\frac{1}{1-u} = 1 + u + u^2 + u^3 + \cdots , \tag{53}$$

which leads to

$$f(z) = \sum_{m\geq 0} \int_\gamma (1-\Phi'_y(z,y))\Phi^m(z,y)\,\frac{dy|}{y^m}. \tag{54}$$

Now Cauchy's coefficient formula,

$$[y^{m-1}]\,g(y) = \frac{1}{\pi}\int_{0+} g(y)\,\frac{dy}{y^m},$$

when applied to the integral in Eq. (54) provides

$$f(z) = \sum_{m\geq 0} [y^{m-1}]\,(1-\Phi'_y(z,y))\Phi^m(z,y), \tag{55}$$

which is the form given in (50).

There only remains to complete the analytic part of the argument. First, by the assumptions made regarding Φ, the point $(0,0)$ is an ordinary point of the algebraic curve $y - \Phi(z,y) = 0$. Thus all branches of the curve are "well separated" from the branch corresponding to $f(z)$. We can thus find $\rho_1 > 0$ and $r_1 > 0$ so that, when z lies in the domain $|z| \leq \rho_1$, we have $|f(z)| \leq r_1$ while $|f_j(z)| > r_1$ for all the conjugate branches $f_j(z) \neq f(z)$. In other words, the use of the integral formula (52) is justified provided we take $|z| < \rho_1$ together with the contour $\gamma = \{y \,/\, |y| = r\}$ for any $r \leq r_1$. We shall henceforth assume that such a choice has been made.

The next condition to be satisfied is the validity of expansion (53) used to derive (54). This requires the inequality $|u| < 1$ when $u = \frac{1}{y}\Phi(z,y)$. should have $|u| < 1$. But the conditions on $\Phi(z,y)$ at $(0,0)$ imply that in a sufficiently small neighbourhood of the origin, i.e. $|z| < \rho_2$ and $|y| < r_2$, the bound,

$$|\Phi(z,y)| \leq K(|z| + |zy| + |y^2|),$$

holds for some positive constant K. Thus, for $|z| < \rho_2$ and $|y| < r_2$, the condition $|\Phi(z,y)| < |y|$ is granted if

$$|z| < |y|\frac{1-|y|}{1+|y|}. \tag{56}$$

We can now conclude the argument. Choose as contour γ a circle centered at the origin with radius $r = \min(r_1, r_2)$. To guarantee condition (56), impose that z be such that

$$|z| < \rho \qquad \text{where} \qquad \rho = \min\left(\rho_1, \rho_2, r\frac{1-r}{1+r}\right).$$

Therefore, Equation (55) holds true. □

Part (ii) of the theorem generalizes the Lagrange inversion formula that corresponds to the "separable" case $\Phi(z,y) = z\phi(y)$. As an example, the coefficients of

$$f(z) = z + z^2 f^2(z) + z^3 f^3(z),$$

admit the "nice" form

$$f_n = \sum_{m\geq 1} \binom{m-1}{n-m+1,\, 5m-3n-2,\, 2n-3m+1}.$$

In general, if Φ comprises p monomials, the formula obtained is a $(p-2)$-fold summation.

5. Analysis of algebraic functions

Algebraic functions can only have a type of singularity constrained to be a branch point. The local expansion at such a singularity is a fractional power series known as a Newton–Puiseux expansion. Singularity analysis is systematically applicable to algebraic functions—hence the characteristic form of asymptotic expansions that involve terms of the form $\omega^n n^{p/q}$ (for some algebraic number ω and some rational exponent p/q). In this section, we develop such basic structural results (Subsection 5.1). However, coming up with effective solutions (i.e., decision procedures) is not obvious in the algebraic case. Hence, a number of nontrivial algorithms are also described (built on top of elimination by resultants or Groebner bases) in order to locate and analyse singularities (Newton's polygon method), and eventually determine the asymptotic form of coefficients. In particular, the multivalued character of algebraic functions creates a need to solve "connection problems". Finally, like in the rational case, positive systems (Subsection 5.2) enjoy special properties that further constrain what can be observed as regards asymptotic behaviours and properties of random structures. Our presentation of positive systems is based on an essential result of the theory, the Drmota–Lalley–Woods theorem, that plays for algebraic functions a rôle quite similar to that of Perron-Frobenius theory for rational functions.

5.1. General algebraic functions. Let $P(z,y)$ be an irreducible polynomial of $\mathbb{C}[z,y]$,

$$P(z,y) = p_0(z)y^d + p_1(z)y^{d-1} + \cdots + p_d(z).$$

The solutions of the polynomial equation $P(z,y) = 0$ define a locus of points (z,y) in $\mathbb{C} \times \mathbb{C}$ that is known as a complex algebraic curve. Let d be the y-degree of P. Then, for each z there are at most d possible values of y. In fact, there exist d values of y "almost always", that is except for a finite number of cases:

— If z_0 is such that $p_0(z_0) = 0$, then there is a reduction in the degree in y and hence a reduction in the number of y-solutions for the particular value of $z = z_0$. One can conveniently regard the points that disappear as "points at infinity".

— If z_0 is such that $P(z_0, y)$ has a multiple root, then some of the values of y will coalesce.

Define the *exceptional set* of P as the set (**R** is the resultant):

$$\Xi[P] := \{z \mid R(z) = 0\}, \ R(z) := \mathbf{R}(P(z,y), \partial_y P(z,y), y).$$

(The quantity $R(z)$ is also known as the discriminant of $P(z,y)$ taken as a function of y.) If $z \notin \Xi[p]$, then we have a guarantee that there exist d distinct solutions to $P(z,y) = 0$, since $p_0(z) \neq 0$ and $\partial_y P(z,y) \neq 0$. Then, by the implicit function theorem, each of the solutions y_j lifts into a locally analytic function $y_j(z)$. What we call a *branch* of the algebraic curve $P(z,y) = 0$ is the choice of such an $y_j(z)$ together with a connected region of the complex plane throughout which this particular $y_j(z)$ is analytic.

EXERCISE 27. Verify that, when z approaches z_0 with $p(z_0) = 0$, then a number at least 1 of the values of y satisfying $P(z,y) = 0$ tend to infinity.

Singularities of an algebraic function can thus only occur if z lies in the exceptional set $\Xi[P]$. At a point z_0 such that $p_0(z_0) = 0$, some of the branches escape to infinity, thereby ceasing to be analytic. At a point z_0 where the resultant polynomial $R(z)$ vanishes but $p_0(z) \neq 0$, then two or more branches collide. This can be either a multiple point (two or more branches happen to assume the same value, but each one exists as an analytic function around z_0) or a branch point (some of the branches actually cease to be analytic).

An example of an exceptional point that is not a branch point is provided by the classical lemniscate of Bernoulli: at the origin, two branches meet while each one is analytic there (see Figure 11).

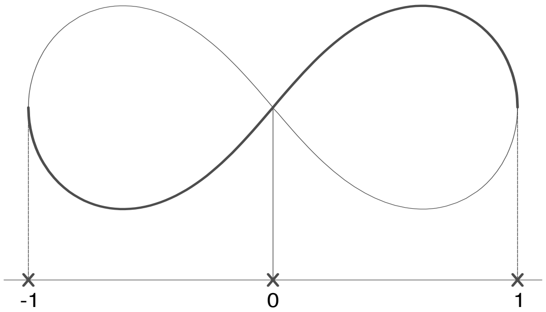


FIGURE 11. The lemniscate of Bernoulli defined by $P(z,y) = (z^2 + y^2)^2 - (z^2 - y^2) = 0$: the origin is a double point where two analytic branches meet.

A partial impression of the topology of a complex algebraic curve may be gotten by first looking at its restriction to the reals. Consider the polynomial equation $P(z,y) = 0$, where

$$P(z,y) = y - 1 - zy^2,$$

which defines the OGF of the Catalan numbers. A rendering of the real part of the curve is given in Figure 12. The complex aspect of the curve as given by $\Im(y)$ as a function of z is also displayed there. In accordance with earlier observations, there are normally two sheets (branches) above each each point. The exceptional set is given by the roots of the discriminant,

$$\mathcal{R} = z(1 - 4z).$$

For $z = 0$, one of the branches escapes at infinity, while for $z = 1/4$, the two branches meet and there is a branch point; see Figure 12.

In summary the exceptional set provides a set of *possible candidates* for the singularities of an algebraic function. This discussion is summarized by the slightly more general lemma that follows.

LEMMA 8.3 (Location of algebraic singularities). *Let $Y(z) \in \mathbb{C}^{alg}[\![z]\!]$ satisfy a polynomial equation $P(z,Y) = 0$. Then, $Y(z)$ is analytic at the origin, i.e., it has a non zero radius of convergence. In addition, it can be analytically continued along any half-line emanating from the origin that does not cross any point of the exceptional set.*

(The fact that an algebraic *series* cannot have radius of convergence equal to 0, i.e., be purely divergent, follows from the method of majorizing series [**52**, II, p 94] as well as from the detailed discussion given below.)

Nature of singularities. We start the discussion with an exceptional point that is placed at the origin (by a translation $z \mapsto z + z_0$) and assume that the equation $P(0,y) = 0$ has k equal roots $y_1, \ldots, y_k$ where $y = 0$ is this common value (by a translation $y \mapsto y + y_0$ or an inversion $y \mapsto 1/y$, if points at infinity are considered). Consider a punctured disk $|z| < r$ that does not include any other exceptional point relative to P. In the argument that follows, we let $y_1, (z), \ldots, y_k(z)$ be analytic determinations of the root that tend to 0 as $z \to 0$.

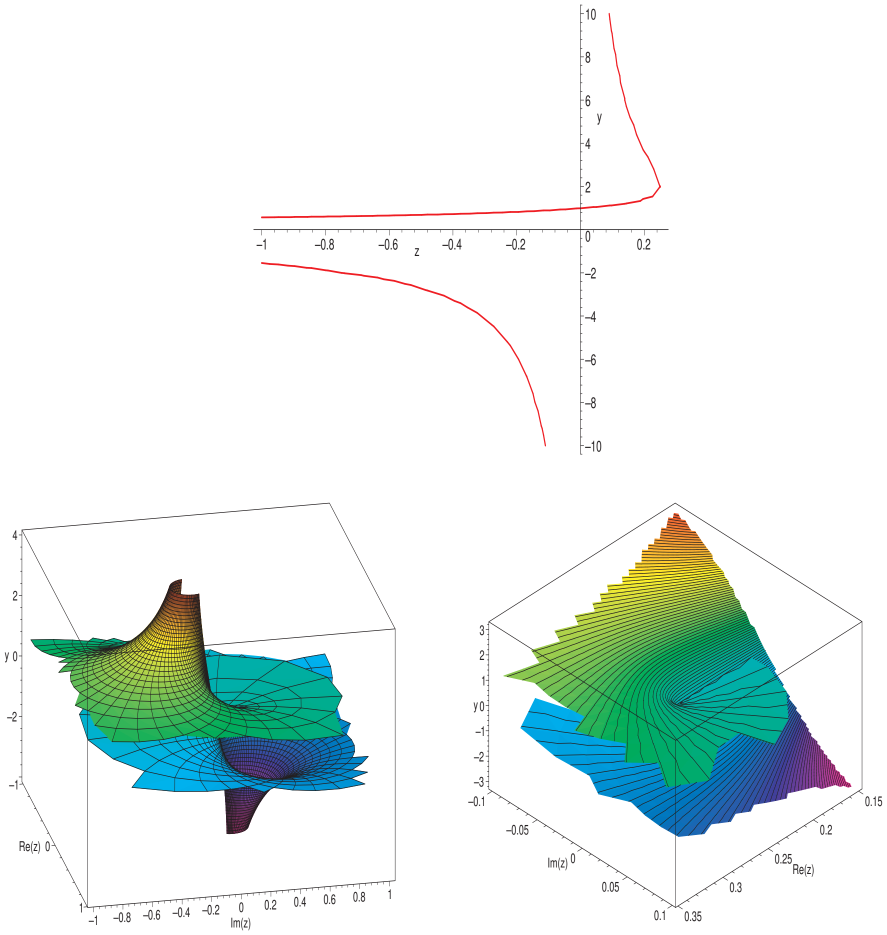


FIGURE 12. The real section of the Catalan curve (top). The complex Catalan curve with a plot of $\Im(y)$ as a function of $z = (\Re(z), \Im(z))$ (bottom left); a blowup of $\Im(y)$ near the branch point at $z = \frac{1}{4}$ (bottom right).

Start at at some arbitrary value interior to the real interval $(0, r)$, where the quantity $y_1(z)$ is locally an analytic function of z. By the implicit function theorem, $y_1(z)$ can be continued analytically along a circuit that starts from z and returns to z while simply encircling the origin (and staying within the punctured disk). Then, by permanence of analytic relations, $y_1(z)$ will be taken into another root, say, $y_1^{(1)}(z)$. By repeating the process, we see that after a certain number of times κ with $1 \le \kappa \le k$, we will have obtained a collection of roots $y_1(z) = y_1^{(0)}(z), \ldots, y_1^{(\kappa)}(z) = y_1(z)$ that form a set of κ distinct values. Such roots are said to form a *cycle*. In this case, $y_1(t^\kappa)$ is an analytic function of t except possibly at 0 where it is continuous and has value 0. Thus, by general principles (regarding removable singularities), it is in fact analytic at 0. This in turn implies the existence of a convergent expansion near 0:

$$y_1(t^\kappa) = \sum_{n=1}^{\infty} c_n t^n. \tag{57}$$

The parameter t is often called the *local uniformizing parameter*, as it reduces a multivalued function to a single value one. This translates back into the world of z: each determination of $z^{1/\kappa}$ yields one of the branches of the multivalued analytic function as

$$y_1(z) = \sum_{n=1}^{\infty} c_n z^{n/\kappa}. \tag{58}$$

Alternatively, with $\omega = e^{2i\pi/\kappa}$ a root of unity, the κ determinations are obtained as

$$y_1^{(j)} = \sum_{n=1}^{\infty} c_n \omega^n z^{n/\kappa},$$

each being valid in a sector of opening $< 2\pi$. (The case $\kappa = 1$ corresponds to an analytic branch.)

If $r = k$, then the cycle accounts for all the roots which tend to 0. Otherwise, we repeat the process with another root and, in this fashion, eventually exhaust all roots. Thus, all the k roots that have value 0 at $z = 0$ are grouped into cycles of size $\kappa_1, \ldots, \kappa_\ell$. Finally, values of y at infinity are brought to zero by means of the change of variables $y = 1/u$, then leading to negative exponents in the expansion of y.

THEOREM 8.11 (Newton–Puiseux expansions at a singularity). *Let $f(z)$ be a branch of an algebraic function $P(z, f(z)) = 0$. In a circular neighbourhood of a singularity ζ slit along a ray emanating from ζ, $f(z)$ admits a fractional series expansion (Puiseux expansion) that is locally convergent and of the form*

$$f(z) = \sum_{k \geq k_0} c_k (z - \zeta)^{k/\kappa},$$

for a fixed determination of $(z - \zeta)^{1/\kappa}$, where $k_0 \in \mathbb{Z}$ and κ is an integer ≥ 2, called the "branching type".

Newton (1643-1727) discovered the algebraic form of Theorem 8.11, published it in his famous treatise *De Methodis Serierum et Fluxionum* (completed in 1671). This method was subsequently developed by Victor Puiseux (1820–1883) so that the name of Puiseux series is customarily attached to fractional series expansions. The argument given above is taken from the neat exposition offered by Hille in [**52**, Ch. 12, vol. II]. It is known as a "monodromy argument", meaning that it consists in following the course (–dromy) of values of a multivalued analytic function along paths in the complex plane till it returns to its original (mono–) value.

Newton polygon. Newton also described a constructive approach to the determination of branching types near a point (z_0, y_0), that by means of the previous discussion can always be taken to be $(0, 0)$. In order to introduce the discussion, let us examine the Catalan generating function near $z_0 = 1/4$. Elementary algebra gives the explicit form of the two branches

$$y_1(z) = \frac{1}{2z}\left(1 - \sqrt{1-4z}\right), \quad y_2(z) = \frac{1}{2z}\left(1 + \sqrt{1-4z}\right),$$

whose forms are consistent with what Theorem 8.11 predicts. If however one starts directly with the equation,

$$P(z, y) \equiv y - 1 - zy^2 = 0$$

then, the translation $z = 1/4 - Z$ (the minus sign is a mere notational convenience), $y = 2 + Y$ yields

$$Q(Z,Y) \equiv -\frac{1}{4}Y^2 + 4Z + 4ZY + ZY^2. \tag{59}$$

Look for solutions of the form $Y = cZ^\alpha(1 + o(1))$ with $c \neq 0$ (the existence is granted *a priori* by the Newton-Puiseux Theorem). Each of the monomials in (59) gives rise to a term of a well determined asymptotic order. respectively $Z^{2\alpha}, Z^1, Z^{\alpha+1}, Z^{2\alpha+1}$. If the equation is to be identically satisfied, then the main asymptotic order of $Q(Z,Y)$ should be 0. Since $c \neq 0$, this can only happen if two or more of the exponents in the sequence $(2\alpha, 1, \alpha+1, 2\alpha+1)$ coincide *and* the coefficients of the corresponding monomial in $P(Z,Y)$ is zero, a condition that is an algebraic constraint on the constant c. Furthermore, exponents of all the remaining monomials have to be larger since by assumption they represent terms of lower asymptotic order.

Examination of all the possible combinations of exponents leads one to discover that the only possible combination arises from the cancellation of the first two terms of Q, namely $-\frac{1}{4}Y^2 + 4Z$, which corresponds to the set of constraints

$$2\alpha = 1, \quad -\frac{1}{4}c^2 + 4 = 0,$$

with the supplementary conditions $\alpha + 1 > 1$ and $2\alpha + 1 > 1$ being satisfied by this choice $\alpha = \frac{1}{2}$. We have thus discovered that $Q(Z,Y) = 0$ is consistent asymptotically with

$$Y \sim 4Z^{1/2}, \quad Y \sim -4Z^{1/2}.$$

The process can be iterated upon subtracting dominant terms. It invariably gives rise to complete formal asymptotic expansions that satisfy $Q(Z,Y) = 0$ (in the Catalan example, these are series in $\pm Z^{1/2}$). Furthermore, elementary majorizations establish that such formal asymptotic solutions represent indeed convergent series. Thus, local expansions of branches have indeed been determined. This is Newton's algorithm for expanding algebraic functions near a branch point.

An algorithmic refinement (also due to Newton) can be superimposed on the previous discussion and is known as the method of *Newton polygons*. Consider a general polynomial

$$Q(Z,Y) = \sum_{j \in J} Z^{a_j} Y^{b_j},$$

and associate to it the finite set of points (a_j, b_j) in $\mathbb{N} \times \mathbb{N}$, which is called the Newton diagram. It is easily verified that the only asymptotic solutions of the form $Y \propto Z^t$ correspond to values of t that are inverse slopes (i.e., $\Delta x/\Delta y$) of lines connecting two or more points of the Newton diagram (this expresses the cancellation condition between two monomials of Q) *and* such that all other points of the diagram are on this line or to the right of it. In other words:

> **Newton's polygon method.** *The possible exponents t such that $Y \sim cZ^t$ is a solution to a polynomial equation correspond to the inverse slopes of the leftmost convex envelope of the Newton diagram. For each viable t, a polynomial equation constrains the possible values of the corresponding coefficient c. Complete expansions are obtained by repeating the process, which means deflating Y from its main term by way of the substitution $Y \mapsto Y - cZ^t$.*

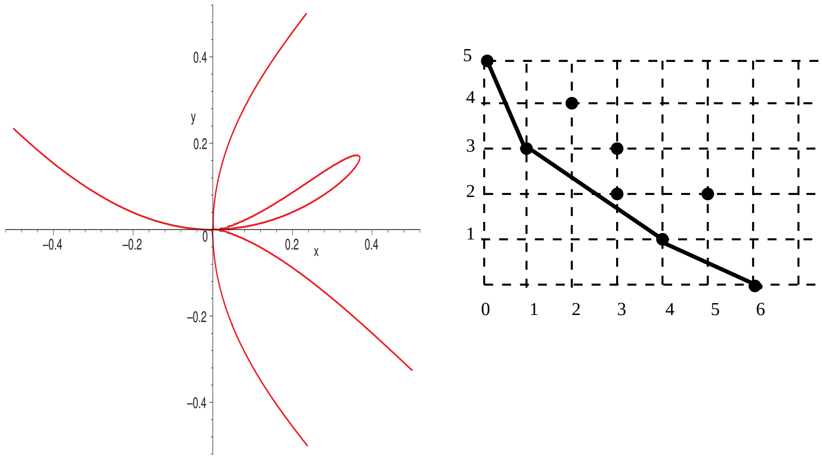


FIGURE 13. The real curve defined by the equation $P = (y-x^2)(y^2-x)(y^2-x^3) - x^3y^3$ near $(0,0)$ (left) and the corresponding Newton diagram (right).

Figure 13 illustrates what goes on in the case of the curve $p = 0$ where

$$\begin{aligned} P(z,y) &= (y-z^2)(y^2-z)(y^2-z^3) - z^3y^3 \\ &= y^5 - y^3z - y^4z^2 + y^2z^3 - 2\,z^3y^3 + z^4y + z^5y^2 - z^6, \end{aligned}$$

considered near the origin. As the partly factored form suggests, we expect the curve to resemble the union of two orthogonal parabolas and of a curve $y = \pm z^{3/2}$ having a cusp, i.e., the union of

$$y = z^2, \quad y = \pm\sqrt{z}, \quad y = \pm z^{3/2},$$

respectively. It is visible on the Newton diagram of the expanded form that the possible exponents $y \propto z^t$ at the origin are the inverse slopes of the segments composing the envelope, that is,

$$t = 2, \quad \frac{1}{2}, \quad \frac{3}{2}.$$

For computational purposes, once determined the branching type κ, the value of k_0 that dictates where the expansion starts, and the first coefficient, the full expansion can be recovered by deflating the function from its first term and repeating the Newton diagram construction. In fact, after a few initial stages of iteration, the method of indeterminate coefficients can always be eventually applied[8]. Computer algebra systems usually have this routine included as one of the standard packages; see [**82**].

EXERCISE 28. Discuss the degree of the algebraic number c in the expansion $Y \sim cZ^t$ in relation to Newton's diagram.

[8]Bruno Salvy, private communication, August 2000

Asymptotic form of coefficients. The Newton–Puiseux theorem describes precisely the local singular structure of an algebraic function. The expansions are valid around a singularity except for one direction. In particular they hold in indented disks of the type required in order to apply the formal translation mechanisms of singularity analysis (Chapter 5).

THEOREM 8.12 (Algebraic asymptotics). *Let $f(z) = \sum_n f_n z^n$ be an algebraic series. Assume that the branch defined by the series at the origin has a unique dominant singularity at $z = \alpha_1$ on its circle of convergence. Then, the coefficient f_n satisfies the asymptotic expansion,*

$$f_n \sim \alpha_1^{-n} \left(\sum_{k \geq k_0} d_k n^{-1-k/\kappa} \right),$$

where $k_0 \in \mathbb{Z}$ and κ is an integer ≥ 2.

If $f(z)$ has several dominant singularities $|\alpha_1| = |\alpha_2| = \cdots = |\alpha_r|$, then there exists an asymptotic decomposition (where ϵ is some small fixed number, $\epsilon > 0$)

$$f_n = \sum_{j=1}^{r} \phi^{(j)}(n) + O((|\alpha_1| + \epsilon))^{-n},$$

where

$$\phi^{(j)}(n) \sim \alpha_j^{-n} \left(\sum_{k \geq k_0^{(j)}} d_k^{(j)} n^{-1-k/\kappa_j} \right),$$

each $k_0^{(j)}$ is in $\mathbb{Z}$, and each κ_j is an integer ≥ 2.

Proof. The directional expansions granted by Theorem 8.11 are of the exact type required by singularity analysis; see Chapter 5. Composite contours should be used in the case of multiple singularities, where each $\phi^{(j)}(n)$ is the contribution obtained by translation of the local singular element. $\square$

In the case of multiple singularities, arithmetic cancellations may occur: consider for instance the case of

$$\frac{1}{\sqrt{1 - \frac{6}{5}z + z^2}} = 1 + 0.60z + 0.04z^2 - 0.36z^3 - 0.408z^4 - \cdots,$$

and refer to the corresponding discussion of rational coefficients, page 11. Fortunately, such delicate situations tend not to arise in combinatorial situations.

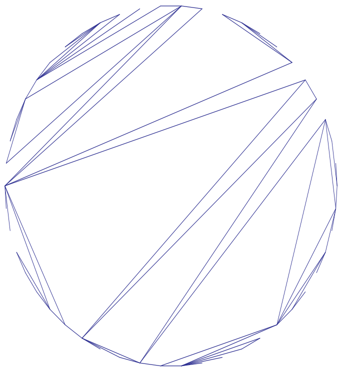
EXAMPLE 17. *Unary-binary trees.* the generating function of unary binary trees is defined by $P(z, f) = 0$ where

$$P(z, y) = y - z - zy - zy^2,$$

so that

$$f(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z} = \frac{1 - z - \sqrt{(1+z)(1-3z)}}{2z}.$$

There exist only two branches: f and its conjugate $\bar{f}$ that form a cycle of size 2 at $\frac{1}{3}$. The singularities of all branches are at $0, -1, \frac{1}{3}$ as is apparent from the explicit form of f or from the defining equation. The branch representing $f(z)$ at the origin is analytic there (by a general argument or by the combinatorial origin of the problem). Thus, the dominant singularity of $f(z)$ is at $\frac{1}{3}$ and it is unique in its modulus class. The "easy" case of Theorem 8.13 then applies once $f(z)$ has been expanded ear $\frac{1}{3}$. As a rule, the organization



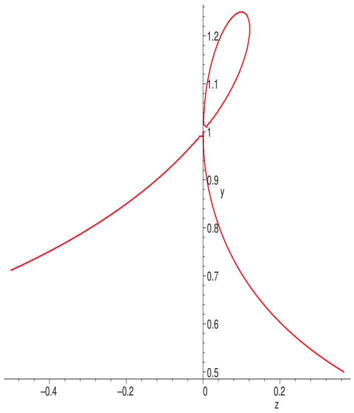


FIGURE 14. Non-crossing graphs: (a) a random connected graph of size 50; (b) the real algebraic curve corresponding to non-crossing forests.

of computations is simpler if one makes use of the local uniformizing parameter with a choice of sign in accordance to the direction along which the singularity is approached. In this case, we set $z = \frac{1}{3} - \delta^2$ and find

$$f(z) = 1 - 3\,\delta + \frac{9}{2}\delta^2 - \frac{63}{8}\delta^3 + \frac{27}{2}\delta^4 - \frac{2997}{128}\delta^5 + \cdots, \qquad \delta = (\frac{1}{3} - z)^{1/2}.$$

This translates immediately into

$$f_n \equiv [z^n]f(z) \sim \frac{3^{n+1/2}}{2\sqrt{\pi n^3}}\left(1 - \frac{15}{16n} + \frac{505}{512n^2} - \frac{8085}{8192n^3} + \cdots\right).$$

The approximation provided by the first three terms is quite good: for $n = 10$ already, it estimates $f_{10} = 835$. with an error less than 1. □

EXERCISE 29. Estimate the growth of the coefficients in the asymptotic expansion of the number of unary-binary trees.

EXAMPLE 18. *Non-crossing forests.* Consider the regular n-gon whose vertices are numbered $1, \ldots, n$. A non-crossing graph (connected graph, tree, forest, etc) is defined as a graph having the property that no two edges edges cross. Let F_n be the number of non-crossing graphs of size n that are forests, i.e., acyclic graphs. It is shown below (Section 6.1; see also [**40**]) that the OGF $F(z)$ satisfies the equation $P(z, F) = 0$, where

$$P(z, y) = y^3 + (z^2 - z - 3)y^2 + (z + 3)y - 1,$$

and that the combinatorial GF starts as

$$F(z) = 1 + 2z + 7z^2 + 33z^3 + 181z^4 + 1083z^5 + \cdots.$$

(This is sequence A054727 of *EIS*.) The exceptional set is mechanically computed as roots of the discriminant

$$R = -z^3(5z^3 - 8z^2 - 32z + 4).$$

Newton's algorithm shows that two of them, say y_0 and y_2, form a cycle of length 2 with $y_0 = 1 - \sqrt{z} + O(z)$, $y_2 = 1 + \sqrt{z} + O(z)$ while it is the "middle branch" $y_1 = 1 + z + O(z^2)$ that corresponds to the combinatorial GF $F(z)$.

The other exceptional points are the roots of the cubic factor of $\mathcal{R}$, namely

$$\Omega = \{-1.93028,\ 0.12158,\ 3.40869\}.$$

Let $\xi \doteq 0.1258$ be the root in $(0, 1)$. By Pringsheim's theorem and the fact that the OGF of an infinite combinatorial class must have a positive dominant singularity in $[0, 1]$, the only possibility for the dominant singularity of $y_1(z)$ is ξ. (For a more general argument, see below.)

For z near ξ, the three branches of the cubic give rise to one branch that is analytic with value approximately 0.67816 and a cycle of two conjugate branches with value near 1.21429 at $z = \xi$. The expansion of the two conjugate branches is of the singular type,

$$\alpha \pm \beta\sqrt{1 - z/\xi},$$

where

$$\alpha = \frac{43}{37} + \frac{18}{37}\xi - \frac{35}{74}\xi^2 \doteq 1.21429, \quad \beta = \frac{1}{37}\sqrt{228 - 981\xi - 5290\xi^2} \doteq 0.14931.$$

The determination with a minus sign must be adopted for representing the combinatorial GF when $z \to \xi^-$ since otherwise one would get negative asymptotic estimates for the nonnegative coefficients. Alternatively, one may examine the way the three real branches along $(0, \xi)$ match with one another at 0 and at ξ^-, then conclude accordingly.

Collecting partial results, we finally get by singularity analysis the estimate

$$F_n = \frac{\beta}{2\sqrt{\pi n^3}}\omega^n \left(1 + O(\frac{1}{n})\right), \qquad \omega = \frac{1}{\xi} \doteq 8.22469$$

where the cubic algebraic number ξ and the sextic β are as above. □

The example above illustrates several important points in the analysis of coefficients of algebraic functions when there are no simple explicit radical forms. First of all a given combinatorial problem determines a unique branch of an algebraic curve at the origin. Next, the dominant singularity has to be identified by "connecting" the combinatorial branch with the branches at every possible singularity of the curve. Finally, computations tend to take place over algebraic numbers and not simply rational numbers.

So far, examples have illustrated the common situation where the exponent at the dominant singularity is $\frac{1}{2}$, which is reflected by a factor of $n^{-3/2}$ in the asymptotic form of coefficients. Our last example shows a case where the exponent assumes a different value, namely $\frac{1}{4}$.

EXAMPLE 19. *"Supertrees"*. Consider the equation $P(z, S) = 0$ where

$$P(z, y) = zy^4 - y^3 + (2\,z + 1)\,y^2 - y + z.$$

The general aspect of the curve is given in Figure 15 and the asymptotic expansion of the branch with positive coefficients, $y(z) = z + z^2 + 3z^3 + 8z^4 + \cdots$, is sought.

The discriminant is found to be

$$\mathcal{R} = z(4z + 3)(4z - 1)^3,$$

so that the dominant singularity of the branch of combinatorial interest is $z = \frac{1}{4}$ where its value equals 1. The translation $z = \frac{1}{2} - Z, y = 1 + Y$ transforms P into

$$\tilde{P}(Z, Y) = (\frac{1}{4} - Z)Y^4 - 4ZY^3 - 8ZY^2 - 8ZY - 4Z.$$

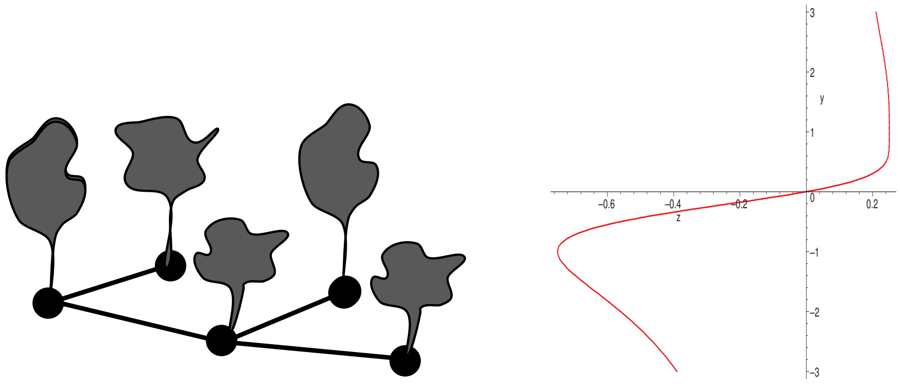


FIGURE 15. Supertrees (left) and their quartic generating function (right).

The main cancellation arises from $\frac{1}{4}Y^4 - 4Z = 0$: this corresponds to a segment of inverse slope $1/4$ in the Newton diagram and accordingly to a cycle formed with 4 conjugate branches. Thus, one has, as $z \to (\frac{1}{4})^-$,

$$y(z) = 1 - 2\sqrt[4]{\frac{1}{4} - z} + \cdots \quad \text{so that} \quad [z^n]y(z) \sim \frac{1}{\sqrt{8}\Gamma(\frac{3}{4})} 4^n\, n^{-5/4},$$

as $n \to \infty$. This exhibits a nonstandard occurrence of $\Gamma(\frac{3}{4})$, of the singular exponent $\frac{1}{4}$ and of the coefficient exponent $-\frac{5}{4}$.

Here is the combinatorial origin of this GF. Consider the GF of complete binary trees,

$$b(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z}.$$

(The function is singular at $\frac{1}{2}$ with value $b(1/2) = 1$.) The composition $b(zb(z))$ represents the GF of "supertrees" (or "trees or trees") obtained by grafting planted binary trees $(zb(z))$ at each node of a complete tree $(b(z))$. The original function $S(z)$ satisfies in fact $S(z^2) = b(zb(z))$. (Naturally, the construction was purposely designed to create an interesting confluence of singularities.) □

Computable coefficient asymptotics. The previous discussion contains the germ of a complete algorithm for deriving an asymptotic expansion of coefficients of any algebraic function. We sketch here the main principles leaving some of the details to the reader. Observe that the problem is a *connection problems*: the "shapes" of the various sheets around each point (including the exceptional points) are known, but it remains to connect them together and see which ones are encountered first when starting from a given branch at the origin.

Algorithm ACA: *Algebraic Coefficient Asymptotics.*
Input: A polynomial $P(z, y)$ with $d = \deg_y P(z, y)$; a series $Y(z)$ such that $P(z, Y) = 0$ and assumed to be specified by sufficiently many initial terms so as to be distinguished from all other branches.
Output: The asymptotic expansion of $[z^n]Y(z)$ whose existence is granted by Theorem 8.12.

The algorithm consists of three main steps: *Preparation*, *Dominant singularities*, and *Translation*.

I. Preparation: Define the discriminant $R(z) = \mathbf{R}(P, P'_y, y)$.

(P_1) Compute the exceptional set $\Xi = \{z \mid R(z) = 0\}$ and the points of infinity $\Xi_0 = \{z \mid p_0(z) = 0\}$, where $p_0(z)$ is the leading coefficient of $P(z, y)$ considered as a function of y.

(P_2) Determine the Puiseux expansions of all the d branches at each of the points of $\Xi \cup \{0\}$ (by Newton diagrams and/or indeterminate coefficients). This includes the expansion of analytic branches as well. Let $\{y_{\alpha,j}(z)\}_{j=1}^{d}$ be the collection of all such expansions at some $\alpha \in \Xi \cup \{0\}$.

(P_3) Identify the branch at 0 that corresponds to $Y(z)$.

II. Dominant singularities (Controlled approximate matching of branches). Let $\Xi_1, \Xi_2, \ldots$ be a partition of the elements of $\Xi \cup \{0\}$ sorted according to the increasing values of their modulus: it is assumed that the numbering is such that if $\alpha \in \Xi_i$ and $\beta \in \Xi_j$, then $|\alpha| < |\beta|$ is equivalent to $i < j$. Geometrically, the elements of Ξ have been grouped in concentric circles. First, a preparation step is needed.

(D_1) Determine a nonzero lower bound δ on the radius of convergence of any local Puiseux expansion of any branch at any point of Ξ. Such a bound can be constructed from the minimal distance between elements of Ξ and from the degree d of the equation.

The sets Ξ_j are to be examined in sequence until it is detected that one of them contains a singularity. At step j, let $\sigma_1, \sigma_2, \ldots, \sigma_s$ be an arbitrary listing of the elements of Ξ_j. The problem is to determine whether any σ_k is a singularity and, in that event, to find the right branch to which it is associated. This part of the algorithm proceeds by controlled numerical approximations of branches and constructive bounds on the minimum separation distance between distinct branches.

(D_2) For each candidate singularity σ_k, with $k \geq 2$, set $\zeta_k = \sigma_k(1 - \delta/2)$. By assumption, each ζ_k is in the domain of convergence of $Y(z)$ and of any $y_{\sigma_k,j}$.

(D_3) Compute a nonzero lower bound η_k on the minimum distance between two roots of $P(\zeta_k, y) = 0$. This separation bound can be obtained from resultant computations.

(D_4) Estimate $Y(\zeta_k)$ and each $y_{\sigma_k,j}(\zeta_k)$ to an accuracy better than $\eta_k/4$. If two elements, $Y(z)$ and $y_{\sigma_k,j}(z)$ are (numerically) found to be at a distance less than η_k for $z = \zeta_k$, then they are matched: σ_k *is* a singularity and the corresponding $y_{\sigma_k,j}$ is the corresponding singular element. Otherwise, σ_k is declared to be a regular point for $Y(z)$ and discarded as candidate singularity.

The main loop on j is repeated until a singularity has been detected., when $j = j_0$, say. The radius of convergence ρ is then equal to the common modulus of elements of Ξ_{j_0}; the corresponding singular elements are retained.

III. Coefficient expansion. Collect the singular elements at all the points σ determined to be a dominant singularity at Phase III. Translate termwise using the singularity analysis rule,

$$(\sigma - z)^{p/\kappa} \mapsto \sigma^{p/\kappa - n} \frac{\Gamma(-p/\kappa + n)}{\Gamma(-p/\kappa)\Gamma(n+1)},$$

and reorganize into descending powers of n, if needed.

This algorithm vindicates the following assertion:

PROPOSITION 8.5 (Decidability of algebraic connections.). *The dominant singularities of a branch of an algebraic function can be determined by the algorithm* ACA *in a finite number of operations in the algebraic closure of the base field,* $\mathbb{C}$ *or* $\overline{\mathbb{Q}}$*.*

5.2. Positive functions and positive systems. The discussion of algebraic singularities specializes nicely to the case of positive functions. We first indicate a procedure that determines the radius of convergence of any algebraic series with *positive* coefficients. The procedure takes advantage of Pringsheim's theorem that allows us to restrict attention to candidate singularities on the positive half-line. It represents a shortcut that is often suitable for human calculation and, in fact, the it systematizes some of the techniques already used implicitly in earlier examples.

Algorithm ROCPAF: *Radius of Convergence of Positive Algebraic Functions.*
Input: A polynomial $P(z, y)$ with $d = \deg_y P(z, y)$; a series $Y(z)$ such that $P(z, Y) = 0$ that is known to have only nonnegative coefficients ($[z^n]Y(z) \geq 0$) and is assumed to be specified by sufficiently many initial terms.
Output: The radius of convergence ρ of $Y(z)$.

Plane-sweep. Let Ξ^+ be the subset of those elements of the exceptional set Ξ which are positive real.

(R_1) Sort the subset of those branches $\{y_{0,j}\}$ at 0^+ that have totally real coefficients. This is essentially a lexicographic sort that only needs the initial parts of each expansion. Set initially $\xi_0 = 0$ and $U(z) = Y(z)$.

(R_2) Sweep over all $\xi \in \Xi^+$ in increasing order. To detect whether a candidate ξ is the dominant positive singularity, proceed as follows:
- — Sort the branches $\{y_{\xi,j}\}$ at ξ^- that have totally real coefficients.
- — using the orders at ξ_0^+ and ξ^-, match the branch $U(z)$ with its corresponding branch at ξ^-, say $V(z)$; this makes use of the total ordering between real branches at ξ_0^+ and ξ^-. If the branch $V(z)$ is singular, then return $\rho = \xi$ as the radius of convergence of $Y(z)$ and use $V(z)$ as the singular element of $Y(z)$ at $z = \rho$; otherwise continue with the next value of $\xi \in \Xi^+$ while replacing $U(z)$ by $V(z)$ and ξ_0 by ξ.

This algorithm is a plane-sweep that takes advantage of the fact that the real branches near a point can be totally ordered; finding the ordering only requires inspection of a finite number of coefficients. The plane-sweep algorithm enables us to trace at each stage the original branch and keep a record of its order amongst all branches. The method works since no two real branches can cross at a point other than a multiple point, such a point being covered as an element of Ξ^+.

We now turn to positive systems. Most of the combinatorial classes known to admit algebraic generating functions involve singular exponents that are multiples of $\frac{1}{2}$. This empirical observation is supported by the fact, to be proved below, that a wide class of positive systems have solutions with a square-root singularity. Interestingly enough, the corresponding theorem is due to independent research by several authors: Drmota [**31**] developed a version of the theorem in the course of studies relative to limit laws in various families of trees defined by context-free grammars; Woods [**100**], motivated by questions of Boolean complexity and finite model theory, gave a form expressed in terms of colouring rules for trees; finally, Lalley [**63**] came across a similarly general result when quantifying return probabilities for random walks on groups. The statement that follows is a fundamental result in the analysis of algebraic systems arising from combinatorics and is (rightly) called the "Drmota-Lalley-Woods" theorem. Notice that the authors of [**31, 63, 100**] prove more: Drmota and Lalley show how to pull out limit Gaussian laws for simple parameters (e.g., as in [**31**] by a perturbative analysis; see Chapter 9); Woods shows how to deduce estimates of coefficients even in some periodic or non-irreducible cases (see definitions below).

In the treatment that follows we start from a polynomial system of equations,

$$\{y_j = \Phi_j(z, y_1, \ldots, y_m)\}, \quad j = 1, \ldots, m.$$

We shall discuss in the next section a class of combinatorial specifications, the "context-free" specifications, that leads systematically to such fixed-point systems. The case of linear systems has been already dealt with, so that we limit ourselves here to *nonlinear systems* defined by the fact that at least one polynomial Φ_j is nonlinear in some of the indeterminates $y_1, \ldots, y_m$.

First, for combinatorial reasons, we define several possible attributes of a polynomial system.

— *Algebraic positivity* (or a-positivity). A polynomial system is said to be *a-positive* if all the component polynomials Φ_j have nonnegative coefficients.

Next, we want to restrict consideration to systems that determine a unique solution vector $(y_1, \ldots, y_m) \in (\mathbb{C}[[z]])^m$. (This discussion is related to 0-dimensionality in the sense alluded to earlier.) Define the *z-valuation* $\mathrm{v}al(\vec{y})$ of a vector $\vec{y} \in \mathbb{C}[[z]]^m$ as the minimum over all j's of the individual valuations[9] $\mathrm{v}al(y_j)$. The distance between two vectors is defined as usual by $d(\vec{y}, \vec{y}') = 2^{-\mathrm{v}al(\vec{y}-\vec{y}')}$. Then, one has:

— *Algebraic properness* (or a-properness). A polynomial system is said to be *a-proper* if it satisfies a Lipschitz condition

$$d(\Phi(\vec{y}), \Phi(\vec{y}\,')) < K d(\vec{y}, \vec{y}\,') \quad \text{for some } K < 1.$$

In that case, the transformation Φ is a contraction on the complete metric space of formal power series and, by the general fixed point theorem, the equation $y = \Phi(y)$ admits a unique solution. In passing, this solution may be obtained by the iterative scheme,

$$\vec{y}^{(0)} = (0, \ldots, 0)^t, \quad \vec{y}^{(h+1)} = \Phi(y^{(h)}), \quad y = \lim_{h\to\infty} y^{(h)}.$$

The key notion is irreducibility. To a polynomial system, $\vec{y} = \Phi(\vec{y})$, associate its *dependency graph* defined as a graph whose vertices are the numbers $1, \ldots, m$ and the edges ending at a vertex j are $k \to j$, if y_j figures in a monomial of $\Phi_k(j)$. (This notion is reminiscent of the one already introduced for linear system on page 8.5.)

— *Algebraic irreducibility* (or a-irreducibility). A polynomial system is said to be *a-irreducible* if its dependency graph is strongly connected.

Finally, one needs a technical notion of periodicity to dispose of cases like

$$y(z) = \frac{1}{2z}\left(1 - \sqrt{1-4z}\right) = z + z^3 + 2z^5 + \cdots,$$

(the OGF of complete binary trees) where coefficients are only nonzero for certain residue classes of their index.

— *Algebraic aperiodicity* (or a-aperiodicity). A power series is said to be *aperiodic* if it contains three monomials (with nonzero coefficients), $z^{e_1}, z^{e_2}, z^{e_3}$, such that $e_2 - e_1$ and $e_3 - e_1$ are relatively prime. A proper polynomial system is said to be aperiodic if each of its component solutions y_j is aperiodic.

THEOREM 8.13 (Positive polynomial systems). *Consider a nonlinear polynomial system $\vec{y} = \Phi(\vec{y})$ that is a-proper, a-positive, and a-irreducible. In that case, all component*

[9]Let $f = \sum_{n=\beta}^{\infty} f_n z^n$ with $f_\beta \neq 0$; the valuation of f is by definition $\mathrm{v}al(f) = \beta$.

solutions y_j have the same radius of convergence $\rho < \infty$. Then, there exist functions h_j analytic at the origin such that

$$y_j = h_j\left(\sqrt{1 - z/\rho}\right) \qquad (z \to \rho^-). \tag{60}$$

In addition, all other dominant singularities are of the form $\rho\omega$ with ω a root of unity. If furthermore the system is a-aperiodic, all y_j have ρ as unique dominant singularity. In that case, the coefficients admit a complete asymptotic expansion of the form

$$[z^n]y_j(z) \sim \rho^{-n}\left(\sum_{k\geq 1} d_k n^{-1-k/2}\right). \tag{61}$$

Proof. The proof consists in gathering by stages consequences of the assumptions. It is essentially based on close examination of "failures" of the implicit function theorem and the way these lead to singularities.

(a) As a preliminary observation, we note that each component solution y_j is an algebraic function that has a nonzero radius of convergence. In particular, singularities are constrained to be of the algebraic type with local expansions in accordance with the Newton-Puiseux theorem (Theorem 8.11).

(b) Properness together with the positivity of the system implies that each $y_j(z)$ has nonnegative coefficients in its expansion at 0, since it is a formal limit of approximants that have nonnegative coefficients. In particular, each power series y_j has a certain nonzero radius of convergence ρ_j. Also, by positivity, ρ_j is a singularity of y_j (by virtue of Pringsheim's theorem). From the nature of singularities of algebraic functions, there exists some order $R \geq 0$ such that each Rth derivative $\partial_z^R y_j(z)$ becomes infinite as $z \to \rho_j^-$.

We establish now that $\rho_1 = \cdots = \rho_m$. In effect, differentiation of the equations composing the system implies that a derivative of arbitrary order r, $\partial_z^r y_j(z)$, is a linear form in other derivatives $\partial_z^r y_j(z)$ of the same order (and a polynomial form in lower order derivatives); also the linear combination and the polynomial form have nonnegative coefficients. Assume a contrario that the radii were not all equal, say $\rho_1 = \cdots = \rho_s$, with the other radii $\rho_{s+1}, \ldots$ being strictly greater. Consider the system differentiated a sufficiently large number of times, R. Then, as $z \to \rho_1$, we must have $\partial_z^R y_j$ tending to infinity for $j \leq s$. On the other hand, the quantities y_{s+1}, etc., being analytic, their Rth derivatives that are analytic as well must tend to finite limits. In other words, because of the irreducibility assumption (and again positivity), infinity *has to* propagate and we have reached a contradiction. Thus, all the y_j have the same radius of convergence and we let ρ denote this common value.

(c_1) The key step consists in establishing the existence of a square-root singularity at the common singularity ρ. Consider first the scalar case, that is

$$y - \phi(z, y) = 0, \tag{62}$$

where ϕ is assumed to depend nonlinearly on y and have nonnegative coefficients. The requirement of properness means that z is a factor of all monomials, except the constant term $\phi(0,0)$.

Let $y(z)$ be the unique branch of the algebraic function that is analytic at 0. Comparison of the asymptotic orders in y inside the equality $y = \phi(z, y)$ shows that (by nonlinearity) we cannot have $y \to \infty$ when z tends to a finite limit. Let now ρ be the radius of convergence of $y(z)$. This argument shows that $y(z)$ is necessarily finite at its singularity ρ. We set $\tau = y(\rho)$ and note that, by continuity $\tau - \phi(\rho, \tau) = 0$.

By the implicit function theorem, a solution (z_0, y_0) of (62) can be continued analytically as $(z, y_0(z))$ in the vicinity of z_0 as long as the derivative with respect to y,

$$J(z_0, y_0) := 1 - \phi'_y(z_0, y_0)$$

remains nonzero. The quantity ρ being a singularity, we must have $J(\rho, \tau) = 0$. (In passing, the system

$$\tau - \phi(\rho, \tau) = 0, \qquad J(\rho, \tau) = 0,$$

determines only finitely many candidates for ρ.) On the other hand, the second derivative $-\phi''_{yy}$ is nonzero at (ρ, τ) (by positivity, since no cancellation can occur); there results by the classical argument on local failures of the implicit function theorem that $y(z)$ has a singularity of the square-root type (see also Chapters 4 and 5). More precisely, the local expansion of the defining equation (62) at (ρ, τ) binds (z, y) locally by

$$-(z - \rho)\phi'_z(\rho, \tau) - \frac{1}{2}(y - \tau)^2 \phi''_{yy}(\rho, \tau) + \cdots = 0,$$

where the subsequent terms are negligible by Newton's polygon method. Thus, we have

$$y - \tau = -\sqrt{\frac{\phi_z(\rho, \tau)}{\phi''_{yy}(\rho, \tau)}}(\rho - z)^{1/2} + \cdots ,$$

the negative determination of the square-root being chosen to comply with the fact that $y(z)$ increases as $z \to \rho^-$. This proves the first part of the assertion in the scalar case.

(c_2) In the multivariate case, we graft an ingenious argument [**63**] that is based on a linearized version of the system to which Perron-Frobenius theory is applicable. First, irreducibility implies that any component solution y_j depends nonlinearly on itself (by possibly iterating Φ), so that a discrepancy in asymptotic behaviours would result for the implicitly defined y_j in the event that some y_j tends to infinity.

Now, the multivariate version of the implicit function theorem grants locally the analytic continuation of any solution $y_1, y_2, \ldots, y_m$ at z_0 provided there is no vanishing of the Jacobian determinant

$$J(z_0, y_1, \ldots, y_m) := \det\left(\delta_{i,j} - \frac{\partial}{\partial y_j}\Phi_i(z_0, y_1, \ldots, y_m)\right),$$

where $\delta_{i,j}$ is Kronecker's symbol. Thus, we must have

$$J(\rho, \tau_1, \ldots, \tau_m) = 0 \qquad \text{where} \quad \tau_j := y_j(\rho).$$

The next argument (we follow Lalley [**63**]) uses Perron-Frobenius theory and linear algebra. Consider the modified Jacobian matrix

$$K(z_0, y_1, \ldots, y_m) := \left(\frac{\partial}{\partial y_j}\Phi_i(z_0, y_1, \ldots, y_m)\right),$$

which represents the "linear part" of Φ. For $z, y_1, \ldots, y_m$ all nonnegative, the matrix K has positive entries (by positivity of Φ) so that it is amenable to Perron-Frobenius theory. In particular it has a positive eigenvalue $\lambda(z, y_1, \ldots, y_m)$ that dominates all the other in modulus. The quantity

$$\widehat{\lambda}(z) = \lambda(y_1(z), \ldots, y_m(z))$$

is increasing as it is an increasing function of the matrix entries that themselves increase with z for $z \geq 0$.

We propose to prove that $\widehat{\lambda}(\rho) = 1$, In effect, $\widehat{\lambda}(\rho) < 1$ is excluded since otherwise $(I - K)$ would be invertible at $z = \rho$ and this would imply $J \neq 0$, thereby contradicting the singular character of the $y_j(z)$ at ρ. Assume *a contrario* $\widehat{\lambda}(\rho) > 1$ in order to exclude the other case. Then, by the increasing property, there would exists $\rho_1 < \rho$ such that $\widehat{\lambda}(\rho_1) = 1$. Let v_1 be a left eigenvector of $K(\rho_1, y_1(\rho_1), \ldots, y_m(\rho_1))$ corresponding to the eigenvalue $\widehat{\lambda}(\rho_1)$. Perron-Frobenius theory grants that such a vector v_1 has all its coefficients that are positive. Then, upon multiplying on the left by v_1 the column vectors corresponding to y and $\Phi(y)$ (which are equal), one gets an identity; this derived identity upon expanding near ρ_1 gives

$$A(z-\rho_1) = -\sum_{i,j} B_{i,j}(y_i(z) - y_i(\rho_1))(y_j(z) - y_j(\rho_1)) + \cdots, \tag{63}$$

where $\cdots$ hides lower order terms and the coefficients $A, B_{i,j}$ are nonnegative with $A > 0$. There is a contradiction in the orders of growth if each y_i is assumed to be analytic at ρ_1 since the left side of (63) is of exact order $(z - \rho_1)$ while the right side is at least as small as to $(z - \rho_1)^2$. Thus, we must have $\widehat{\lambda}(\rho) = 1$ and $\widehat{\lambda}(x) < 1$ for $x \in (0, \rho)$.

A calculation similar to (63) but with ρ_1 replaced by ρ shows finally that, if

$$y_i(z) - y_i(\rho) \sim \gamma_i(\rho - z)^{\alpha},$$

then consistency of asymptotic expansions implies $2\alpha = 1$, that is $\alpha = \frac{1}{2}$. (The argument here is similar to the first stage of a Newton polygon construction.) We have thus proved that the component solutions $y_j(z)$ have a square-root singularity. (The existence of a complete expansion in powers of $(\rho - z)^{1/2}$ results from examination of the Newton diagram.) The proof of the general case (60) is at last completed.

(d) In the aperiodic case, we first observe that each $y_j(z)$ cannot assume an infinite value on its circle of convergence $|z| = \rho$, since this would contradict the boundedness of $|y_j(z)|$ in the open disk $|z| < \rho$ (where $y_j(\rho)$ serves as an upperbound). Consequently, by singularity analysis, the Taylor coefficients of any $y_j(z)$ are $O(n^{-1-\eta})$ for some $\eta > 1$ and the series representing y_j at the origin converges on $|z| = \rho$.

For the rest of the argument, we observe that if $y = \Phi(z, \vec{y})$, then $y = \Phi^{\langle m \rangle}(z, \vec{y})$ where the superscript denotes iteration of the transformation Φ in the variables $\vec{y} = (y_1, \ldots, y_m)$. By irreducibility, $\Phi^{\langle m \rangle}$ is such that *each* of its component polynomials involves *all* the variables.

Assume that there would exists a singularity ρ^* of some $y_j(z)$ on $|z| = \rho$. The triangular inequality yields $|y_j(\rho^*)| < y_j(\rho)$ where strictness is related to the general aperiodicity argument encountered at several other places in this book. But then, the modified Jacobian matrix $K^{\langle m \rangle}$ of $\Phi^{\langle m \rangle}$ taken at the $y_j(\rho^*)$ has entries dominated strictly by the entries of $K^{\langle m \rangle}$ taken at the $y_j(\rho)$. There results (see page 17) that the dominant eigenvalue of $K^{\langle m \rangle}(z, \vec{y}_j(\rho^*))$ must be strictly less than 1. But this would imply that $I - K^{\langle m \rangle}(z, \vec{y}_j(\rho^*))$ is intervertible so that the $y_j(z)$ would be analytic at ρ^*. A contradiction has been reached: ρ is the sole dominant singularity of each y_j and this concludes the argument. $\square$

We observe that the dominant singularity is obtained amongst the positive solutions of the system

$$\vec{\tau} = \Phi(\rho, \vec{\tau}), \quad J(\rho, \vec{\tau}) = 0.$$

For the Catalan GF, this yields

$$\tau - 1 - \rho\tau^2 = 0, \quad 1 - 2\rho\tau = 0,$$

giving back (as expected) $\rho = \frac{1}{4}$, $\tau = \frac{1}{2}$. For three coloured trees (with y_1, y_2, y_3 representing A, B, C), the system is formed of the specialization of the defining equations (40), namely,

$$\tau_1 - \rho - (\tau_2 + \tau_3)^2 = 0, \quad \tau_2 - (\tau_3 + \tau_1)^2 = 0, \quad \tau_3 - (\tau_1 + \tau_2)^2 = 0.$$

together with the Jacobian condition

$$\det \begin{pmatrix} 1 & -2\tau_2 - 2\tau_3 & -2\tau_3 - 2\tau_2 \\ -2\tau_1 - 2\tau_3 & 1 & -2\tau_3 - 2\tau_1 \\ -2\tau_1 - 2\tau_2 & -2\tau_2 - 2\tau_1 & 1 \end{pmatrix} = 0.$$

It is found (by elimination) that

$$\rho = \frac{1}{4}\alpha(3\alpha + 1), \quad \text{where} \quad 4\alpha^3 + 4\alpha^2 + \alpha - 1 = 0,$$

with $\alpha \doteq 0.34681$ and $\rho \doteq 0.177681$ being the only feasible solution to the constraints. Thus, the number of 3-coloured trees grows roughly like $5.62^n n^{-3/2}$.

EXERCISE 30. Match the computation of the dominant singularity of 3-coloured trees against the determination of the minimal polynomial $R(z, A)$ of the previous section.

6. Combinatorial applications of algebraic functions

In this section, we first present context-free specifications that admit a direct translation into polynomial systems (Section 6.1). When particularized to formal languages, this gives rise to context-free languages (Section 6.2) that, provided an unambiguity condition is met, lead to algebraic generating functions. An important subclass, especially as regards computer science applications, is that of simple families of trees succinctly presented in Section 6.3.

The next two subsections introduce objects whose constructions still lead to algebraic functions, but in a non-obvious way. This includes: walks with a finite number of allowed basic jumps (Section 6.4) and planar maps (Section 6.5). In that case, bivariate functional equations are induced by the combinatorial decompositions. The common form is

$$\Phi(z, u, F(z, u), h_1(z), \ldots, h_r(z)) = 0, \tag{64}$$

where Φ is a known polynomial and the unknowns are F and $h_1, \ldots, h_r$. Specific methods are to be appealed to in order to attain solutions to such functional equations that would seem at first glance to be grossly underdetermined. Random walks lead to a linear version of (64) that is treated by the so-called "kernel method". Maps lead to nonlinear versions that are solved by means of Tutte's "quadratic method". In both cases, the strategy consists in binding z and u by forcing them to lie on an algebraic curve (suitably chosen in order to eliminate the dependency on $F(z, u)$), and then pulling out the algebraic consequences of such a specialization.

6.1. Context-free specifications. A *context-free system* is a collection of combinatorial equations,

$$\begin{cases} \mathcal{C}_1 & = & \Phi_1(\vec{a}, \mathcal{C}_1, \ldots, \mathcal{C}_m) \\ \vdots & & \vdots \\ \mathcal{C}_m & = & \Phi_m(\vec{a}, \mathcal{C}_1, \ldots, \mathcal{C}_m), \end{cases} \tag{65}$$

where $\vec{a} = (a_1, \ldots)$ is a vector of atoms and each of the Φ_j only involves the combinatorial constructions of disjoint union and cartesian product. A combinatorial class $\mathcal{C}$ is said to be context-free if it is definable as the first component ($\mathcal{C} = \mathcal{C}_1$) of a well-founded context-free system. The terminology comes from linguistics and it stresses the fact that objects can be "freely" generated by the rules in (65), this without any constraints imposed by an outside context[10].

For instance the class of plane binary trees defined by

$$\mathcal{B} = e + (i \times \mathcal{B} \times \mathcal{B}) \qquad (e, i \text{ atoms})$$

is a context-free class. The class of general plane trees defined by

$$\mathcal{G} = o \times \mathtt{sequence}(\mathcal{G}) \qquad (o \text{ an atom})$$

is definable by the system

$$\mathcal{G} = o \times \mathcal{F}, \quad \mathcal{F} = \mathbf{1} + (\mathcal{F} \times \mathcal{G}),$$

with $\mathcal{F}$ defining forests, and so it is also context-free. (This example shows more generally that sequences can always be reduced to polynomial form.)

Context-free specifications may be used to describe all sorts of combinatorial objects. For instance, the class $\mathcal{T}$ of triangulations of convex polygons is specified symbolically by

$$\mathcal{T} = \nabla + (\nabla \times \mathcal{T}) + (\mathcal{T} \times \nabla) + (\mathcal{T} \times \nabla \times \mathcal{T}), \tag{66}$$

where ∇ represents a generic triangle.

The general symbolic rules given in Chapter 1 apply in all such cases. Therefore the Drmota-Lalley-Woods theorem (Theorem 8.13) provides the asymptotic solution to an important category of problems.

PROPOSITION 8.6 (Context-free specifications). *A context-free class $\mathcal{C}$ admits an OGF that satisfies a polynomial system obtained from the specification by the translation rules:*

$$\mathcal{A} + \mathcal{B} \mapsto A + B, \quad \mathcal{A} \times \mathcal{B} \mapsto A \cdot B.$$

The OGF $C(z)$ is an algebraic function to which algebraic asymptotics applies. In particular, a context-free class $\mathcal{C}$ that gives rise to an algebraically aperiodic irreducible system has an enumeration sequence satisfying

$$C_n \sim \frac{\gamma}{\sqrt{\pi n^3}} \omega^n,$$

where γ, ω are computable algebraic numbers.

This last result explains the frequently encountered estimates involving a factor of $n^{-3/2}$ (corresponding to a square-root singularity of the OGF) that can be found throughout analytic combinatorics.

EXERCISE 31. If $\mathcal{A}, \mathcal{B}$ are context-free specifications then: (i) the sequence class $\mathcal{C} = \mathtt{sequence}(\mathcal{A})$ is context-free; (ii) the substitution class $\mathcal{D} = \mathcal{A}[b \mapsto \mathcal{B}]$ is context-free.

We detail below an example from combinatorial geometry.

[10] Formal language theory also defines context-sensitive grammars where each rule (called a production) is applied only if it is enabled by some external context. Context-sensitive grammars have greater expressive power than context-free ones, but they depart significantly from decomposability since they are surrounded by strong undecidability properties; accordingly context-sensitive grammars cannot be associated with any global generating function formalism.

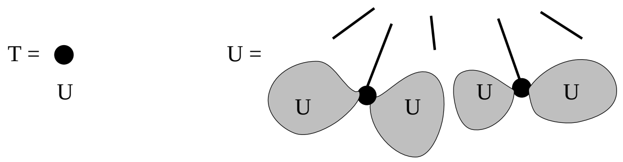
EXAMPLE 20. *Planar non-crossing configurations.* The enumeration of non-crossing planar configurations is discussed here at some level of generality. (An analytic problem in this orbit has been already treated in Example 18.) The purpose is to illustrate the fact that context-free descriptions can model naturally very diverse sorts of objects including particular topological-geometric configurations. The problems considered have their origin in combinatorial musings of the Rev. T.P. Kirkman in 1857 and were revisited in 1974 by Domb and Barett [**30**] for the purpose of investigating certain perturbative expansions of statistical physics. Our presentation follows closely the synthesis offered in [**40**].

Consider for each value of n the regular n-gon built from vertices taken for convenience to be the n complex roots of unity and numbered $0, \ldots, n-1$. A non-crossing graph is a graph on this set of vertices such that no two of its edges cross. From there, one defines non-crossing connected graphs, non-crossing forests (that are acyclic), and non-crossing trees (that are acyclic and connected); see Figure 16. Note that there is a well-defined orientation of the complex plane and also that the various graphs considered can always be rooted in some canonical way (e.g., on the vertex of smallest index) since the placement of vertices is rigidly fixed.

Trees. A non-crossing tree is rooted at 0. To the root vertex, is attached an ordered collection of vertices, each of which has an end-node ν that is the common root of two non-crossing trees, one on the left of the edge $(0, \nu)$ the other on the right of $(0, \nu)$. Let $\mathcal{T}$ denote the class of trees and $\mathcal{U}$ denote the class of trees whose root has been severed. With o denoting a generic node, we then have

$$\mathcal{T} = o \times \mathcal{U}, \quad \mathcal{U} = \mathtt{sequence}(\mathcal{U} \times o \times \mathcal{U}),$$

which corresponds graphically to the "butterfly decomposition":



In terms of OGF, this gives the system

$$\{T = zU,\, U = (1 - zU^2)^{-1}\} \quad \Longleftrightarrow \quad \{T = zU,\, U = 1 + UV,\, V = zU^2\}, \tag{67}$$

where the latter form corresponds to the expansion of the sequence operator. Consequently, T satisfies $T = T^3 - zT + z^2$, which by Lagrange inversion gives $T_n = \frac{1}{2n-1}\binom{3n-3}{n-1}$.

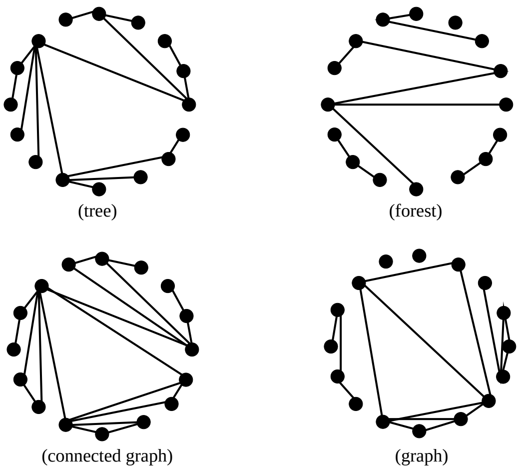
Forests. A (non-crossing) forest is a non-crossing graph that is acyclic. In the present context, it is not possible to express forests simply as sequences as trees, because of the geometry of the problem.

Starting conventionally from the root vertex 0 and following all connected edges defines a "backbone" tree. To the left of every vertex of the tree, a forest may be placed. There results the decomposition (expressed directly in terms of OGF's),

$$F = 1 + T[z \mapsto zF], \tag{68}$$

where T is the OGF of trees and F is the OGF of forests. In (68), the term $T[z \mapsto zF]$ denotes a functional composition. A context-free specification in standard form results mechanically from (67) upon replacing z by zF, namely

$$F = 1 + T,\; T = zFU,\; U = 1 + UV,\; V = zFU^2. \tag{69}$$



Configuration / OGF	Coefficients (exact / asymptotic)
Trees (*EIS*: **A001764**)	$z + z^2 + 3z^3 + 12z^4 + 55z^5 + \cdots$
$T^3 - zT + z^2 = 0$	$\frac{1}{2n-1}\binom{3n-3}{n-1}$
	$\sim \frac{\sqrt{3}}{27\sqrt{\pi n^3}}(\frac{27}{4})^n$
Forests (*EIS*: **A054727**)	$1 + z + 2z^2 + 7z^3 + 33z^4 + 181z^5 \cdots$
$F^3 + (z^2 - z - 3)F^2 + (z+3)F - 1 = 0$	$\sum_{j=1}^{n} \frac{1}{2n-j}\binom{n}{j-1}\binom{3n-2j-1}{n-j}$
	$\sim \frac{0.07465}{\sqrt{\pi n^3}}\,(8.22469)^n$
Connected graphs (*EIS*: **A007297**)	$z + z^2 + 4z^3 + 23z^4 + 156z^5 + \cdots$
$C^3 + C^2 - 3zC + 2z^2 = 0$	$\frac{1}{n-1}\sum_{j=n-1}^{2n-3}\binom{3n-3}{n+j}\binom{j-1}{j-n+1}$
	$\sim \frac{2\sqrt{6}-3\sqrt{2}}{18\sqrt{\pi n^3}}\left(6\sqrt{3}\right)^n$
Graphs (*EIS*: **A054726**)	$1 + z + 2z^2 + 8z^3 + 48z^4 + 352z^5 + \cdots$
$G^2 + (2z^2 - 3z - 2)G + 3z + 1 = 0$	$\frac{1}{n}\sum_{j=0}^{n-1}(-1)^j\binom{n}{j}\binom{2n-2-j}{n-1-j}2^{n-1-j}$
	$\sim \frac{\sqrt{140-99\sqrt{2}}}{4\sqrt{\pi n^3}}\left(6+4\sqrt{2}\right)^n$

FIGURE 16. (Top) Non-crossing graphs: a tree, a forest, a connected graph, and a general graph. (Bottom) The enumeration of non-crossing configurations by algebraic functions.

This system is irreducible and aperiodic, so that the asymptotic shape of F_n is of the form $\gamma\omega_n n^{-3/2}$, as predicted by Proposition 8.6. This agrees with the precise formula determined in Example 18.

Graphs. Similar constructions (see [**40**]) give the OGF's of connected graphs and general graphs. The results are summarized in Figure 16. Note the common shape of the asymptotic estimates and also the fact that binomial expressions are always available in accordance with Theorem 8.10. ☐

Note on "tree-like" structures. A context-free specification can always be regarded as defining a class of trees. Indeed, if the jth term in the construction Φ_j is "coloured" with the pair (i, j), it is seen that a context-free system yields a class of trees whose nodes are tagged by pairs (i, j) in a way that is consistent with the system's rules (65). However, despite this correspondence, it is often convenient to preserve the possibility of operating directly with objects[11] when the tree aspect is unnatural. By a terminology borrowed from the theory of syntax analysis in computer science, such trees are referred to as "parse trees" or "syntax trees".

EXERCISE 32. The parse trees associated mechanically with the specification of triangulations above are in bijective correspondence with binary (rooted plane) trees.

6.2. Context-free languages. Let $\mathcal{A}$ be a fixed finite alphabet whose elements are called letters. A *grammar* G is a collection of equations

$$(70) \qquad G \quad : \quad \begin{cases} \mathcal{L}_1 & = \quad \Psi_1(\vec{a}, \mathcal{L}_1, \ldots, \mathcal{L}_m) \\ \vdots & \qquad \vdots \\ \mathcal{L}_m & = \quad \Psi_m(\vec{a}, \mathcal{L}_1, \ldots, \mathcal{L}_m), \end{cases}$$

where each Ψ_j involves only the operations of union ($\cup$) and catenation product ($\cdot$) with $\vec{a}$ the vector of letters in $\mathcal{A}$. For instance,

$$\Psi_1(\vec{a}, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3) = a_2 \cdot \mathcal{L}_2 \cdot \mathcal{L}_3 \cup a_3 \cup \mathcal{L}_3 \cdot a_2 \cdot \mathcal{L}_1.$$

A solution to (70) is an m-tuple of languages over the alphabet $\mathcal{A}$ that satisfies the system. By convention, one declares that the grammar G defines the first component, $\mathcal{L}_1$.

To each grammar (70), one can associate a context-free specification (65) by transforming unions into disjoint union, '$\cup$' $\mapsto$ '$+$', and catenation into cartesian products, '$\cdot$' $\mapsto$ '$\times$'. Let $\widehat{G}$ be the specification associated in this way to the grammar G. The objects described by $\widehat{G}$ appear in this perspective to be trees (see the discussion above regarding parse trees). Let h be the transformation from trees of $\widehat{G}$ to languages of G that lists letters in infix (i.e., left-to-right) order: we call such an h the erasing transformation since it "forgets" all the structural information contained in the parse tree and only preserves the succession of letters. Clearly, application of h to the combinatorial specifications determined by $\widehat{G}$ yields languages that obey the grammar G. For a grammar G and a word $w \in \mathcal{A}^\star$, the number of parse trees $t \in \widehat{G}$ such that $h(t) = w$ is called the *ambiguity coefficient* of w with respect to the grammar G; this quantity is denoted by $\kappa_G(w)$.

A grammar G is unambiguous if all the corresponding ambiguity coefficients are either 0 or 1. This means that there is a bijection between parse trees of $\widehat{G}$ and words of

[11] Some authors have even developed a notion of "object grammars"; see for instance [**32**] itself inspired by techniques of polyomino surgery in [**29**].

the language described by G: each word generated is uniquely "parsable" according to the grammar. From Proposition 8.6, we have immediately:

PROPOSITION 8.7. *Given a context-free grammar G, the ordinary generating function of the language $L_G(z)$, counting words* with *multiplicity, is an algebraic function. In particular, a context-free language that admits an unambiguous grammar specification has an ordinary generating function $L(z)$ that is an algebraic function.*

This theorem originates from early works of Chomsky and Schützenberger [**22**] which have exerted a strong influence on the philosophy of the present book.

For example consider the Łukasiewicz language

$$\mathcal{L} = (a \cdot \mathcal{L} \cdot \mathcal{L} \cdot \mathcal{L}) \cup b.$$

This can be interpreted as the set of functional terms built from the ternary symbol a and the nullary symbol b:

$$\begin{aligned} \mathcal{L} &= \{b,\ abbb,\ aabbbbb,\ ababbbb, \ldots\} \\ &\simeq \{b,\ a(b,b,b),\ a(a(b,b,b),b,b),\ a(b,a(b,b,b),b), \ldots\}, \end{aligned}$$

where $\simeq$ denotes combinatorial isomorphism. It is easily seen that the terms are in bijective correspondence with their parse trees, themselves isomorphic to ternary trees. Thus the grammar is unambiguous, so that the OGF equation translates directly from the grammar,

$$L(z) = zL(z)^3 + z. \tag{71}$$

As another example, we revisit Dyck paths that are definable by the grammar,

$$D = 1 \cup (a \cdot D \cdot b \cdot D)\,, \tag{72}$$

where a denotes ascents and b denotes descents. Each word in the language must start with a letter a that has a unique matching letter b and thus it is uniquely parsable according to the grammar (72). Since the grammar is unambiguous, the OGF reads off:

$$D(z) = z + z^2 D(z)^2.$$

EXERCISE 33. Investigate the relations between parse trees of Łukasiewicz words (71) and of non-crossing trees.

EXERCISE 34. Extend the discussion to trees where node degrees are constrained to be multiples of some $d \geq 2$. Relate combinatorially this problem to d-ary trees.

6.3. Simple families of trees. Meir and Moon in a classic paper [**69**] were the first to discover the possibility of general asymptotic results concerning simple families of trees. Recall that a *simple family of trees* is defined as the class of all rooted unlabelled plane trees such that the outdegrees of nodes are constrained to belong to a finite set $\Omega \in \mathbb{N}$. The degree polynomial is

$$\phi(y) := \sum_{\omega \in \Omega} c_\omega u^\omega,$$

where c_ω is a positive "multiplicity coefficient" (in the case of pure sets Ω, one has $c_\omega = 1$; the situation $c_\omega \in \mathbb{N}$ covers trees with c_ω allowed colours for a node of degree ω). Then, the OGF of all trees in the family satisfies

$$T(z) = z\phi(T(z)).$$

This is simply a scalar system to which the preceding theory applies.

Assume Ω to be aperiodic, in the sense that the common gcd of the elements of Ω equals 1. Then the pair (ρ, τ) (where ρ is the radius of convergence of $T(z)$ and $\tau = T(\rho)$) is a solution to

$$\tau - \rho\phi(\tau) = 0, \qquad 1 - \rho\phi'(\tau) = 0$$

implying the "characteristic equation",

$$\phi(\tau) - \tau\phi'(\tau) = 0. \tag{73}$$

It can be seen that there is a unique positive solution to (73), which determines

$$\rho = \frac{\tau}{\phi(\tau)} = \frac{1}{\phi'(\tau)}.$$

Then, one finds in agreement with Theorems 8.12 and and 8.13,

$$[z^n]T(z) \sim \gamma \frac{\rho^{-n}}{\sqrt{\pi n^3}}, \quad \gamma = (2\phi(\tau)\phi''(\tau))^{-1/2}.$$

The result extends to finite weighted multisets of allowable node degrees (hence to the nonplanar labelled case). Periodicities are also easily reduced to the general case. See the last chapter of this book for a summary.

EXERCISE 35. Examine the enumeration of "semi-simple" families of trees that are binary trees defined by the fact that edges have size between two bounds $c < d$, binary nodes have size 0, and end-nodes have size between two bounds $a < b$. Such trees are relevant to the enumeration of secondary structures of nucleic sequences in biology [**89**].

EXERCISE 36. Examine the enumeration of plane trees that are coloured in r different ways, where the colours at each node are constrained to satisfy a finite set of compatibility rules [**100**].

EXERCISE 37. A branching process conditioned by fixing the size n of its total progeny leads to a simple family of trees.

EXERCISE 38. Show that non-plane trees with node degrees restricted to some finite Ω yield generating functions that are never algebraic but that the type of the dominant singularity is still a square-root. [See Pólya and Otter's works [**75, 79**].]

6.4. Walks and the kernel method. Start with a set Ω that is a finite subset of $\mathbb{Z}$ and is called the set of *jumps*. A *walk* (relative to Ω) is a sequence $w = (w_0, w_1, \ldots, w_n)$ such that $w_0 = 0$ and $w_{i+1} - w_i \in \Omega$, for all i, $0 \le i < n$. A *nonnegative walk* satisfies $w_i \ge 0$ and an *excursion* is a nonnegative walk such that, additionally, $w_n = 0$. The quantity n is called the length of the walk or the excursion. For instance, Dyck paths and Motzkin paths analysed in Section 3.5 are excursions that correspond to $\Omega = \{-1, +1\}$ and $\Omega = \{-1, 0, +1\}$ respectively. (Walks and excursions can be viewed as particular cases of paths in a graph in the sense of Section 3.3, with the graph taken to be the infinite set $\mathbb{Z}_{>0}$ of integers.)

We propose to determine f_n, the number of excursions of length n and type Ω, via the corresponding OGF

$$F(z) = \sum_{n=0}^{\infty} f_n z^n.$$

In fact, we shall determine the more general BGF

$$F(z,u) := \sum_{n,k} f_{n,k} u^k z^n,$$

where $f_{n,k}$ is the number of walks of length n and final altitude k (*i.e.*, the value of w_n in the definition of a walk is constrained to equal k). In particular, one has $F(z) = F(z,0)$.

We let $-c$ denote the smallest (negative) value of a jump, and d denote the largest (positive) jump. A fundamental rôle is played in this discussion by the "characteristic polynomial" of the walk,

$$S(y) := \sum_{\omega\in\Omega} y^\omega = \sum_{j=-c}^{d} S_j y^j$$

that is a Laurent polynomial[12]. Observe that the bivariate generating function of generalized walks where intermediate values are allowed to be negative, with z marking the length and u marking the final altitude, is rational:

$$G(z,u) = \frac{1}{1 - zS(u)}. \tag{74}$$

Returning to nonnegative walks, the main result to be proved below is the following: *For each finite set $\Omega \in \mathbb{Z}$, the generating function of excursions is an algebraic function that is explicitly computable from Ω.* There are many ways to view this result. The problem is usually treated within probability theory by means of Wiener-Hopf factorizations [**81**]. In contrast, Labelle and Yeh [**61**] show that an unambiguous context-free specification can be systematically constructed, a fact that is sufficient to ensure the algebraicity of the GF $F(z)$. (Their approach is based implicitly on the construction of a finite pushdown automaton itself equivalent, by general principles, to a context-free grammar.) The Labelle-Yeh construction reduces the problem to a large, but somewhat "blind", combinatorial preprocessing, and, for analysts it has the disadvantage of not extracting a simpler (and noncombinatorial) structure inherent in the problem. The method described below is often known as the "kernel" method. It takes its inspiration from exercises in the 1968 edition of Knuth's book [**58**] (Ex. 2.2.1.4 and 2.2.1.11) where a new approach was proposed to the enumeration of Catalan and Schroeder objects. The technique has since been extended and systematized by several authors; see for instance [**5, 6, 15, 33, 34**].

Let $f_n(u) = [z^n]F(z,u)$ be the generating function of walks of length n with u recording the final altitude. There is a simple recurrence relating $f_{n+1}(u)$ to $f_n(u)$, namely,

$$f_{n+1}(u) = S(u)\cdot f_n(u) - r_n(u), \tag{75}$$

where $r_n(u)$ is a Laurent polynomial consisting of the sum of all the monomials of $S(u)f_n(u)$ that involve negative powers[13] of u:

$$r_n(u) := \sum_{j=-c}^{-1} u^j \left([u^j]\, S(u) f_n(u)\right) = \{u^{<0}\} S(u) f_n(u). \tag{76}$$

[12] If Ω is a set, then the coefficients of S lie in $\{0,1\}$. The treatment above applies in all generality to cases where the coefficients are arbitrary positive real numbers. This accounts for probabilistic situations as well as multisets of jump values.

[13] The convenient notation $\{u^{<0}\}$ denotes the singular part of a Laurent expansion: $\{u^{<0}\}f(z) := \sum_{j<0} \left([u^j]f(u)\right)\cdot u^j$.

The idea behind the formula is to subtract the effect of those steps that would take the walk below the horizontal axis. For instance, one has

$$\begin{array}{lll} S(u) = \dfrac{S_{-1}}{u} + O(1) & : & r_n(u) = \dfrac{S_{-1}}{u} f_n(0) \\ S(u) = \dfrac{S_{-2}}{u^2} + \dfrac{S_{-1}}{u} + O(1) & : & r_n(u) = \left(\dfrac{S_{-2}}{u^2} + \dfrac{S_{-1}}{u}\right) f_n(0) + \dfrac{S_{-2}}{u} f_n'(0) \end{array}$$

and generally:

$$\lambda_j(u) = \frac{1}{j!}\{u^{<0}\} u^j S(u). \tag{77}$$

Thus, from (75) and (76) (multiply by z^{n+1} and sum), the generating function $F(z,u)$ satisfies the fundamental functional equation

$$F(z,u) = 1 + zS(u)F(z,u) - z\{u^{<0}\}\left(S(u)F(z,u)\right). \tag{78}$$

Explicitly, one has

$$F(z,u) = 1 + zS(u)F(z,u) - z\sum_{j=0}^{c-1} \lambda_j(u) \left[\frac{\partial^j}{\partial u^j} F(z,u)\right]_{u=0}, \tag{79}$$

for Laurent polynomials $\lambda_j(u)$ that depend on $S(u)$ in an effective way by (77).

The main equations (78) and (79) involve one unknown bivariate GF, $F(z,u)$ and c univariate GF's, the partial derivatives of F specialized at $u = 0$. It is true, but not at all obvious, that the single functional equation (79) fully determines the $c+1$ unknowns. The basic technique is known as "cancelling the kernel" and it relies on strong analyticity properties; see the book by Fayolle *et al.* [**34**] for deep ramifications. The form of (79) to be employed for this purpose starts by grouping on one side the terms involving $F(z,u)$,

$$F(z,u)(1 - zS(u)) = 1 - z\sum_{j=0}^{c-1} \lambda_j(u) G_j(z), \qquad G_j(z) := \left[\frac{\partial^j}{\partial u^j} F(z,u)\right]. \tag{80}$$

If the right side was not present, then the solution would reduce to (74). In the case at hand, from the combinatorial origin of the problem and implied bounds, the quantity $F(z,u)$ is bivariate analytic at $(z,u) = (0,0)$ (by elementary exponential majorizations on the coefficients). The main principle of the kernel method consists in *coupling* the values of z and u in such a way that $1 - zS(u) = 0$, so that $F(z,u)$ disappears from the picture. A condition is that both z and u should remain small (so that F remains analytic). Relations between the partial derivatives are then obtained from such a specializations, $(z,u) \mapsto (z, u(z))$, which happen to be just in the right quantity.

Consequently, we consider the "kernel equation",

$$1 - zS(u) = 0, \tag{81}$$

which is rewritten as

$$u^c = z \cdot (u^c S(u)).$$

Under this form, it is clear that the kernel equation (81) defines $c + d$ branches of an algebraic function. A local analysis (Newton's polygon method) shows that, amongst these $c+d$ branches, there are c branches that tend to 0 as $z \to 0$ while the other d tend to infinity as $z \to 0$. Let $u_0(z), \dots, u_{c-1}(z)$ be the c branches that tend to 0, that we call "small" branches. In addition, we single out $u_0(z)$, the "principal" solution, by the reality condition

$$u_0(z) \sim \gamma z^{1/c}, \quad \gamma := (S_c)^{1/c} \in \mathbb{R}_{>0} \qquad (z \to 0^+).$$

By local uniformization (57), the conjugate branches are given locally by

$$u_\ell(z) = u_0(\mathrm{e}^{2i\ell\pi} z) \qquad (z \to 0^+).$$

Coupling z and u by $u = u_\ell(z)$ produces interesting specializations of Equation (80). In that case, (z, u) is close to $(0, 0)$ where F is bivariate analytic so that the substitution is admissible. By substitution, we get

$$1 - z \sum_{j=0}^{c-1} \lambda_j(u_\ell(z)) \left[\frac{\partial^j}{\partial u^j} F(z, u)\right]_{u=0}, \qquad \ell = 0 \,.\,.\, c-1. \tag{82}$$

This is now a linear system of c equation in c unknowns (the partial derivatives) with algebraic coefficients that, in principle, determines $F(z, 0)$.

A convenient approach to the solution of (82) is due to Mireille Bousquet-Mélou. The argument goes as follows. The quantity

$$M(u) := u^c - zu^c \sum_{j=0}^{c-1} \lambda_j(u) \frac{\partial^j}{\partial u^j} F(z, 0) \tag{83}$$

can be regarded as a polynomial in u. It is monic while it vanishes by construction at the c small branches $u_0, \ldots, u_{c-1}$. Consequently, one has the factorization,

$$M(u) = \prod_{\ell=0}^{c-1} (u - u_\ell(z)). \tag{84}$$

Now, the constant term of $M(u)$ is otherwise known to equal $-zS_{-c}F(z, 0)$, by the definition (83) of $M(u)$ and by Equation (77) specialized to $\lambda_0(u)$. Thus, the comparison of constant terms between (83) and (84) provides us with an explicit form of the OGF of excursions:

$$F(z, 0) = \frac{(-1)^{c-1}}{S_{-c} z} \prod_{\ell=0}^{c-1} u_\ell(z).$$

One can then finally return to the original functional equation and pull the BGF $F(z, u)$. We can thus state:

PROPOSITION 8.8 (Kernel method for walks). *Let Ω be a finite step of jumps and let $S(u)$ be the characteristic polynomial of Ω. In terms of the c small branches of the "kernel" equation,*

$$1 - zS(u) = 0,$$

denoted by $u_0(z), \ldots, u_{c-1}(z)$, the generating function of excursions is expressible as

$$F(z) = \frac{(-1)^{c-1}}{zS_{-c}} \prod_{\ell=0}^{c-1} u_\ell(z) \quad \textit{where } S_{-c} = [u^{-c}]S(u)$$

is the multiplicity (or weight) of the smallest element $-c \in \Omega$.

More generally the bivariate generating function of nonnegative walks is bivariate algebraic and given by

$$F(z, u) = \frac{1}{u^c - z\,(u^c S(u))} \prod_{\ell=0}^{c-1} (u - u_\ell(z)).$$

We give next a few examples illustrating this kernel technique.

Trees and Łukasiewicz codes. A particular class of walks is of special interest; it corresponds to cases where $c = 1$, that is, the largest jump in the negative direction has amplitude 1. Consequently, $\Omega + 1 = \{0, s_1, s_2, \dots, s_d\}$. In that situation, combinatorial theory teaches us the existence of fundamental isomorphisms between walks defined by steps Ω and trees whose degrees are constrained to lie in $1 + \Omega$. The correspondence is by way of Łukasiewicz codes[14] (‘also known as ‘Polish” prefix codes, “Polish” prefix notation), and from it we expect to find tree GF’s in such cases.

As regards generating functions, there now exists only *one* small branch, namely the solution $u_0(z)$ to $u_0(z) = z\phi(u_0(z))$ (where $\phi(u) = uS(u)$) that is analytic at the origin. One then has $F(z) = F(z,0) = \frac{1}{z}u_0(z)$, so that the walk GF is determined by

$$F(z,0) = \frac{1}{z}u_0(z), \quad u_0(z) = z\phi(u_0(z)), \quad \phi(u) := uS(u).$$

This form is consistent with what is already known regarding the enumeration of simple families of trees. In addition, one finds

$$F(z,u) = \frac{1 - u^{-1}u_0(z)}{1 - zS(u)} = \frac{u - u_0(z)}{u - z\phi(u)}.$$

Classical specializations are rederived in this way:

— the Catalan walk (Dyck path), defined by $\Omega = \{-1, +1\}$ and $\phi(u) = 1 + u^2$, has
$$u_0(z) = \frac{1}{2z}\left(1 - \sqrt{1 - 4z^2}\right);$$

— the Motzkin walk, defined by $\Omega = \{-1, 0, +1\}$ and $\phi(u) = 1 + u + u^2$ has
$$u_0(z) = \frac{1}{2z}\left(1 - z - \sqrt{1 - 2z - 3z^2}\right);$$

— the modified Catalan walk, defined by $\Omega = \{-1, 0, 0 + 1\}$ (with two steps of type 0) and $\phi(u) = 1 + 2u + u^2$, has
$$u_0(z) = \frac{1}{2z}\left(1 - 2z - \sqrt{1 - 4z}\right);$$

— the d-ary tree walk (the excursions encode d-ary trees) defined by $\Omega = \{-1, d - 1\}$, has $u_0(z)$ that is defined implicitly by
$$u_0(z) = z(1 + u_0(z)^d).$$

Examples of the general case. Take now $\Omega = \{-2, -1, 1, 2\}$ so that

$$S(u) = u^{-2} + u^{-1} + u + u^2.$$

Then, $u_0(z), u_1(z)$ are the two branches that vanish as $z \to 0$ of the curve

$$y^2 = z(1 + y + y^3 + y^4).$$

The linear system that determines $F(z,0)$ and $F'(z,0)$ is

$$\begin{cases} 1 - \left(\dfrac{z}{u_0(z)^2} + \dfrac{z}{u_0(z)}\right)F(z,0) - \dfrac{z}{u_0(z)}F'(z,0) & = & 0 \\ 1 - \left(\dfrac{z}{u_1(z)^2} + \dfrac{z}{u_1(z)}\right)F(z,0) - \dfrac{z}{u_1(z)}F'(z,0) & = & 0 \end{cases}$$

[14] Such a code [**66**] is obtained by a preorder traversal of the tree, recording a jump of $r - 1$ when a node of outdegree r is encountered. The sequence of jumps gives rise to an excursion followed by an extra -1 jump.

(derivatives are taken with respect to the second argument) and one finds

$$F(z,0) = -\frac{1}{z}u_0(z)u_1(z), \qquad F'(z,0) = \frac{1}{z}(u_0(z)+u_1(z)+u_0(z)u_1(z)).$$

This gives the number of walks, through a combination of series expansions,

$$F(z) = 1 + 2z^2 + 2z^3 + 11z^4 + 24z^5 + 93z^6 + 272z^7 + 971z^8 + 3194z^9 + \cdots .$$

A single algebraic equation for $F(z) = F(z,0)$ is then obtained by elimination (*e.g.*, via Groebner bases) from the system:

$$\begin{cases} u_0^2 - z(1+u_0+u_0^3+u_0^4) &= 0 \\ u_1^2 - z(1+u_1+u_1^3+u_1^4) &= 0 \\ zF + u_0u_1 &= 0 \end{cases}$$

Elimination shows that $F(z)$ is a root of the equation

$$z^4y^4 - z^2(1+2z)y^3 + z(2+3z)y^2 - (1+2z)y + 1 = 0.$$

For walks corresponding to $\Omega = \{-2,-1,0,1,2\}$, we find similarly $F(z) = -\frac{1}{z}u_0(z)u_1(z)$, where u_0, u_1 are the small branches of $y^2 = z(1+y+y^2+y^3+y^4)$, the expansion starts as

$$F(z) = 1 + z + 3z^2 + 9z^3 + 32z^4 + 120z^5 + 473z^6 + 1925z^7 + 8034z^8 + \cdots ,$$

and $F(z)$ is a root of the equation

$$z^4y^4 - z^2(1+z)y^3 + z(2+z)y^2 - (1+z)y + 1 = 0.$$

6.5. Maps and the quadratic method. A (planar) map is a connected planar graph together with an embedding into the plane. In all, generality, loops and multiple edges are allowed. A planar map therefore separates the plane into regions called faces (Figure 17). The maps considered here are in addition rooted, meaning that a face, an incident edge, and an incident vertex are distinguished. In this section, only rooted maps are considered[15]. When representing rooted maps, we shall agree to draw the root edge with an arrow pointing away from the root node, and to take the root face as that face lying to the left of the directed edge (represented in grey on Figure 17).

Tutte launched in the 1960's a large census of planar maps, with the intention of attacking the four-colour problem by enumerative techniques[16]; see [**16, 93, 94, 95, 96**]. There exists in fact an entire zoo of maps defined by various degree or connectivity constraints. In this chapter, we shall limit ourselves to conveying a flavour of this vast theory, with the goal of showing how algebraic functions arise. The presentation takes its inspiration from the book of Goulden and Jackson [**48**, Sec. 2.9]

Let $\mathcal{M}$ be the class of all maps where size is taken to be the number of edges. Let $M(z,u)$ be the BGF of maps with u marking the number of edges on the outside face. The basic surgery performed on maps distinguishes two cases based upon the nature of the

[15]Nothing is lost regarding asymptotic properties of random structures when a rooting is imposed. The reason is that a map has, with probability exponentially close to 1, a trivial automorphism group; consequently, almost all maps of m edges can be rooted in $2m$ ways (by choosing an edge, and an orientation of this edge), and there is an almost uniform $2m$-to-1 correspondence between unrooted maps and rooted ones.

[16]The four-colour theorem to the effect that every planar graph can be coloured using only four colours was eventually proved by Appel and Haken in 1976, using structural graph theory methods supplemented by extensive computer search.

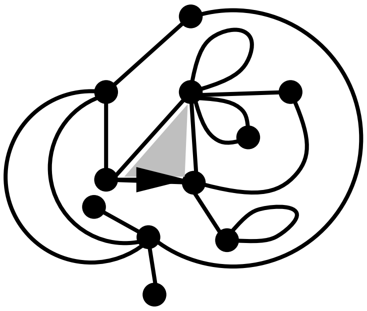


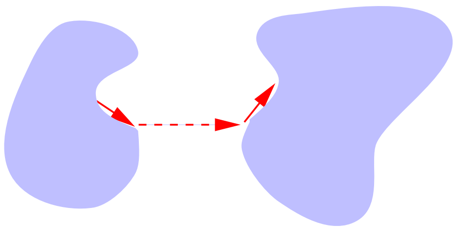
FIGURE 17. A planar map.

root edge. A rooted map will be declared to be isthmic if the root edge r of map μ is an "isthmus" whose deletion would disconnect the graph. Clearly, one has,

$$\mathcal{M} = o + \mathcal{M}^{(i)} + \mathcal{M}^{(n)}, \tag{85}$$

where $\mathcal{M}^{(i)}$ (resp. $\mathcal{M}^{(n)}$) represent the class of isthmic (resp. non-isthmic) maps and 'o' is the graph consisting of a single vertex and no edge. There are accordingly two ways to build maps from smaller ones by adding a new edge.

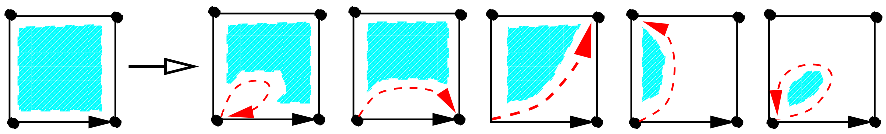
(i) The class of all isthmic maps is constructed by taking two arbitrary maps and joining them together by a new root edge, as shown below:



The effect is to increase the number of edges by 1 (the new root edge) and have the root face degree become 2 (the two sides of the new root edge) plus the sum of the root face degrees of the component maps. The construction is clearly revertible. In other words, the BGF of $\mathcal{M}^{(i)}$ is

$$M^{(i)}(z,u) = zu^2 M(z,u)^2. \tag{86}$$

(ii) The class of non-isthmic maps is obtained by taking an already existing map and adding an edge that preserves its root node and "cuts across" its root face in some unambiguous fashion (so that the construction should be revertible). This operation will therefore result in a new map with an essentially smaller root-face degree. For instance, there are 5 ways to cut across a root face of degree 4, namely,



This corresponds to the linear transformation

$$u^4 \quad \mapsto \quad zu^5 \quad + \quad zu^4 \quad + \quad zu^3 \quad + \quad zu^2 \quad + \quad zu^1.$$

In general the effect on a map with root face of degree k is described by the transformation $u^k \mapsto z(1-u^{k+1})/(1-u)$; equivalently, each monomial $g(u) = u^k$ is transformed into $u(g(1) - ug(u))/(1-u)$. Thus, the OGF of $\mathcal{M}^{(n)}$ involves a discrete difference operator:

$$M^{(n)}(z,u) = zu\frac{M(z,1) - uM(z,u)}{1-u}. \tag{87}$$

Collecting the contributions from (86) and (87) in (85) then yields the basic functional equation,

$$M(z,u) = 1 + u^2 zM(z,u)^2 + uz\frac{M(z,1) - uM(z,u)}{1-z}. \tag{88}$$

The functional equation (88) binds two unknown functions, $M(z,u)$ and $M(z,1)$. Much like in the case of walks, it would seem to be underdetermined. Now, a method due to Tutte and known as the quadratic method provides solutions. Following Tutte and the account in [**48**, p. 138], we consider momentarily the more general equation

$$(g_1 F(z,u) + g_2)^2 = g_3, \tag{89}$$

where $g_j = G_j(z,u,h(z))$ and the G_j are explicit functions—here the unknown functions are $F(z,u)$ and $h(z)$ (cf. $M(z,u)$ and $M(z,1)$ in (88)). Bind u and z in such a way that the left side of (89) vanishes, that is, substitute $u = u(z)$ (a yet unknown function) so that $g_1 F + g_2 = 0$. Since the left-hand side of (89) now has a double root in u, so must the right-hand side, which implies

$$g_3 = 0, \qquad \left.\frac{\partial g_3}{\partial u}\right|_{u=u(z)}. \tag{90}$$

The original equation has become a system of two equations in two unknowns that determines implicitly $h(z)$ and $u(z)$. From there, elimination provides individual equations for $u(z)$ and for $h(z)$. (If needed, $F(z,u)$ can then be recovered by solving a quadratic equation.) It will be recognized that, if the quantities q_1, g_2, g_3 are polynomials, then the process invariably yields solutions that are algebraic functions.

We now carry out this programme in the case of maps and Equation (88). First, isolate $M(z,u)$ by completing the square, giving

$$\left(M(z,u) - \frac{1}{2}\frac{1-u+u^2 z}{u^2 z(1-u)}\right)^2 = Q(z,u) + \frac{M(z,1)}{u(1-u)}, \tag{91}$$

where

$$Q(z,u) = \frac{z^2u^4 - 2zu^2(u-1)(2u-1) + (1-u^2)}{4u^4z^2(1-u)^2}.$$

Next, the condition expressing the existence of a double root is

$$Q(z,u) + \frac{1}{u(1-u)}M(z,1) = 0, \quad Q'_u(z,u) + \frac{2u-1}{u^2(1-u)^2}M(z,1) = 0.$$

It is now easy to eliminate $M(z,1)$, since the dependency in M is linear, and a straightforward calculation shows that $u = u(z)$ should satisfy

$$\left(u^2z + (u-1)\right)\left(u^2z + (u-1)(2u-3)\right) = 0.$$

The first parametrization would lead to $M(z,1) = 1/z$ which is not admissible. Thus, $u(z)$ is to be taken as the root of the second factor, with $M(z,1)$ being defined parametrically by

$$z = \frac{(1-u)(2u-3)}{u^2}, \quad M(z,1) = -u\frac{3u-4}{(2u-3)^2}.$$

The change of parameter $u = 1 - 1/w$ reduces this further to the "Lagrangean form",

$$z = \frac{w}{1-3w}, \quad M(z,1) = \frac{1-4w}{(1-3w)^2}. \tag{92}$$

To this the Lagrange inversion theorem can be applied. The number of maps with n edges, $M_n = [z^n]M(z,1)$ is then determined as

$$M_n = 2\frac{(2n)!3^n}{n!(n+2)!},$$

and one obtains Sequence **A000168** of the *EIS*:

$$M(z,1) = 1 + 2z + 9z^2 + 54z^3 + 378z^4 + 2916z^5 + 24057z^6 + 208494z^7 + \cdots .$$

We refer to [**48**, Sec. 2.9] for detailed calculations (that are nowadays routinely performed with assistance of a computer algebra system). Currently, there exist many applications of the method to maps satisfying all sorts of combinatorial constraints (*e.g.*, multiconnectivity); see [**83**] for a recent panorama.

The derivation above has purposely stressed a parametrized approach as this constitutes a widely applicable approach in many situations. In a simple case like this, we may also eliminate u and solve explicitly for $M(z,1)$, to wit,

$$M(z) \equiv M(z,1) = -\frac{1}{54\,z^2}\left(1 - 18z - (1-12z)^{3/2}\right).$$

It is interesting to note that the singular exponent here is $\frac{3}{2}$, a fact further reflected by the somewhat atypical factor of $n^{-5/2}$ in the asymptotic form of coefficients:

$$M_n \sim \frac{2}{\sqrt{\pi n^5}}12^n \qquad (n \to \infty).$$

Accordingly, randomness properties of maps are appreciably different from what is observed in trees and many commonly encountered context-free objects.

7. Notes

A very detailed discussion of rational functions in combinatorial analysis is given in Stanley's book [**87**] that focusses on algebraic aspects. The main characterizations and closure properties are developed there on a firm algebraic basis. The Perron-Frobenius theory is covered extensively in Gantmacher's reference book on matrix theory [**45**, Ch. 13] as well as in most courses on stochastic processes because of its relevance to finite Markov chains; see for instance the excellent appendix of Karlin and Taylor's course [**54**]. The connections between rational series and formal languages form the subject of the book by Berstel and Reutenauer [**12**] that also discusses rational series in noncommutative indeterminates. Connections between graphs, matrices, and rational functions appear in reference books in algebraic combinatorics, like those by Biggs [**14**] or Godsil [**46**].

The intimate relations between finite automata, regular expressions, and languages belong to the protohistory of computer science in the 1950's. The connection with generating functions evolved largely from joint research [**22**] of the mathematician Marco Schützenberger and the linguist Noam Chomsky in the second half of the 1950's. The method of transfer matrices is closely related to finite automata and it belongs to the folklore of statistical physics; see Temperley's monograph [**91**].

Lattice paths are fundamental objects of combinatorics and several books have been devoted already to their enumerative aspects; see [**71**] for a treatment that offers insights on motivations coming from neighbouring areas of classical statistics and probability theory. Historically, the connection between lattice paths and continued fractions has surfaced independently in works of Touchard (topological configurations [**92**]), I.J. Good (random walks [**47**]), Lenard (one-dimensional statistical mechanics [**65**]), Szekeres (combinatorics of some of Ramanujan's identities), Flajolet (histories and dynamic data structures [**35, 36**]), Jackson (combinatorics of Ising models [**53**]), and Read (chord diagrams [**80**]). The basic results are still rediscovered and published periodically. General syntheses appear in [**35, 37, 48**] and the relation to orthogonal polynomials is well developed in Godsil's book [**46**]. The presentation given here bases itself on reference [**37**], a study motivated more specifically by the formal theory of birth-and-death processes.

Algebraic functions are the modern counterpart of the study of curves by classical Greek mathematicians. They are either approached by algebraic methods (this is the core of algebraic geometry) or by transcendental methods. For our purposes, however, only rudiments of the theory of curves are needed. For this, there exist several excellent introductory books, of which we recommend the ones by Abhyankar [**1**], Fulton [**44**], and Kirwan [**56**]. On the algebraic side, we have striven to provide an introduction to algebraic functions that requires minimal apparatus. At the same time the emphasis has been put somewhat on algorithmic aspects, since most algebraic models are nowadays likely to be treated with the help of computer algebra. As regards symbolic computational aspects, we recommend the treatise by von zur Gathen and Jürgen[**98**] for background, while polynomial systems are excellently reviewed in the book by Cox, Little, and O'Shea [**27**].

In the combinatorial domain, algebraic functions have been used early: in Euler and Segner's enumeration of triangulations (1753) as well as in Schröder's famous "*vier combinatorische Probleme*" described in [**88**, p. 177]. A major advance was the realization by Chomsky and Schützenberger that algebraic functions are the "exact" counterpart of context-free grammars and languages (see again the historic paper [**22**]). A masterful summary of the early theory appears in the proceedings edited by Berstel [**10**] while a modern and precise exposition forms the subject of Chapter 6 of Stanley's book [**88**]. On the analytic-asymptotic side, many researchers have long been aware of the power of Puiseux expansions in conjunction with some version of singularity analysis (often in the form of the Darboux–Pólya method: see [**79**] based on Pólya's classic paper [**78**] of 1937). However, there appeared to be difficulties in coping with the fully general problem of algebraic coefficient asymptotics [**18, 70**]. We believe that Section 5.1 sketches the first complete theory. In the case of positive systems, the "Drmota-Lalley-Woods" theorem is the key to most problems encountered in practice—its importance should be clear from the developments of Section 5.2.

Bibliography

[1] Abhyankar, S.-S. *Algebraic geometry for scientists and engineers*. American Mathematical Society, 1990.

[2] Abramowitz, M., and Stegun, I. A. *Handbook of Mathematical Functions*. Dover, 1973. A reprint of the tenth National Bureau of Standards edition, 1964.

[3] Aho, A. V., and Corasick, M. J. Efficient string matching: an aid to bibliographic search. *Communications of the ACM 18* (1975), 333–340.

[4] Andrews, G. E. *The Theory of Partitions*, vol. 2 of *Encyclopedia of Mathematics and its Applications*. Addison–Wesley, 1976.

[5] Banderier, C. *Combinatoire analytique des chemins et des cartes*. PhD thesis, Université Paris VI, Apr. 2001. In preparation.

[6] Banderier, C., Bousquet-Mélou, M., Denise, A., Flajolet, P., Gardy, D., and Gouyou-Beauchamps, D. On generating functions of generating trees. In *Formal Power Series and Algebraic Combinatorics* (June 1999), C. Martínez, M. Noy, and O. Serra, Eds., Universitat Politècnica de Catalunya, pp. 40–52. (Proceedings of FPSAC'99, Barcelona. Also available as INRIA Res. Rep. 3661, April 1999.).

[7] Barbour, A. D., Holst, L., and Janson, S. *Poisson approximation*. The Clarendon Press Oxford University Press, New York, 1992. Oxford Science Publications.

[8] Bender, E. A. Asymptotic methods in enumeration. *SIAM Review 16*, 4 (Oct. 1974), 485–515.

[9] Bentley, J., and Sedgewick, R. Fast algorithms for sorting and searching strings. In *Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (1997), SIAM Press.

[10] Berstel, J., Ed. *Séries Formelles*. LITP, University of Paris, 1978. (Proceedings of a School, Vieux–Boucau, France, 1977).

[11] Berstel, J., and Perrin, D. *Theory of codes*. Academic Press Inc., Orlando, Fla., 1985.

[12] Berstel, J., and Reutenauer, C. *Les séries rationnelles et leurs langages*. Masson, Paris, 1984.

[13] Bieberbach, L. *Theorie der gewöhnlichen Differentialgleichungen*, vol. LXVI of *Grundlehren der mathematischen Wissenschaften*. Springer Verlag, Berlin, 1953.

[14] Biggs, N. *Algebraic Graph Theory*. Cambridge University Press, 1974.

[15] Bousquet-Mélou, M., and Petkovšek, M. Linear recurrences with constant coefficients: the multivariate case. *Discrete Mathematics ??* (2000), ??? In press.

[16] Brown, W. G., and Tutte, W. T. On the enumeration of rooted non-separable planar maps. *Canadian Journal of Mathematics 16* (1964), 572–577.

[17] Burge, W. H. An analysis of binary search trees formed from sequences of nondistinct keys. *JACM 23*, 3 (July 1976), 451–454.

[18] Canfield, E. R. Remarks on an asymptotic method in combinatorics. *Journal of Combinatorial Theory*, Series A *37* (1984), 348–352.

[19] Cartier, P., and Foata, D. *Problèmes combinatoires de commutation et réarrangements*, vol. 85 of *Lecture Notes in Mathematics*. Springer Verlag, 1969.

[20] Cazals, F. Monomer-dimer tilings. *Studies in Automatic Combinatorics 2* (1997). Electronic publication `http://algo.inria.fr/libraries/autocomb/autocomb.html`.

[21] Chihara, T. S. *An Introduction to Orthogonal Polynomials*. Gordon and Breach, New York, 1978.

[22] Chomsky, N., and Schützenberger, M. P. The algebraic theory of context–free languages. In *Computer Programing and Formal Languages* (1963), P. Braffort and D. Hirschberg, Eds., North Holland, pp. 118–161.

[23] Chyzak, F., and Salvy, B. Non-commutative elimination in Ore algebras proves multivariate identities. *Journal of Symbolic Computation 26*, 2 (1998), 187–227.

[24] Clément, J., Flajolet, P., and Vallée, B. Dynamical sources in information theory: A general analysis of trie structures. *Algorithmica 29*, 1/2 (2001), 307–369.

[25] Comtet, L. Calcul pratique des coefficients de Taylor d'une fonction algébrique. *Enseignement Math. 10* (1964), 267–270.

[26] Comtet, L. *Advanced Combinatorics*. Reidel, Dordrecht, 1974.

[27] Cox, D., Little, J., and O'Shea, D. *Ideals, Varieties, and Algorithms: an Introduction to Computational Algebraic Geometry and Commutative Algebra*, 2nd ed. Springer, 1997.

[28] David, F. N., and Barton, D. E. *Combinatorial Chance*. Charles Griffin, London, 1962.

[29] DELEST, M.-P., AND VIENNOT, G. Algebraic languages and polyominoes enumeration. *Theoretical Computer Science 34* (1984), 169–206.

[30] DOMB, C., AND BARRETT, A. Enumeration of ladder graphs. *Discrete Mathematics 9* (1974), 341–358.

[31] DRMOTA, M. Systems of functional equations. *Random Structures & Algorithms 10*, 1–2 (1997), 103–124.

[32] DUTOUR, I., AND FÉDOU, J.-M. Object grammars and random generation. *Discrete Mathematics and Theoretical Computer Science 2* (1998), 47–61.

[33] FAYOLLE, G., AND IASNOGORODSKI, R. Two coupled processors: the reduction to a Riemann-Hilbert problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 47*, 3 (1979), 325–351.

[34] FAYOLLE, G., IASNOGORODSKI, R., AND MALYSHEV, V. *Random walks in the quarter-plane*. Springer-Verlag, Berlin, 1999.

[35] FLAJOLET, P. Combinatorial aspects of continued fractions. *Discrete Mathematics 32* (1980), 125–161.

[36] FLAJOLET, P., FRANÇON, J., AND VUILLEMIN, J. Sequence of operations analysis for dynamic data structures. *Journal of Algorithms 1* (1980), 111–141.

[37] FLAJOLET, P., AND GUILLEMIN, F. The formal theory of birth-and-death processes, lattice path combinatorics, and continued fractions. *Advances in Applied Probability 32* (2000), 750–778.

[38] FLAJOLET, P., HATZIS, K., NIKOLETSEAS, S., AND SPIRAKIS, P. On the robustness of interconnections in random graphs: A symbolic approach. *Theoretical Computer Science ??* (2001), ???–??? Accepted for publication. A preliminary version appears as INRIA Research Report **4069**, November 2000, 21 pages.

[39] FLAJOLET, P., KIRSCHENHOFER, P., AND TICHY, R. F. Deviations from uniformity in random strings. *Probability Theory and Related Fields 80* (1988), 139–150.

[40] FLAJOLET, P., AND NOY, M. Analytic combinatorics of non-crossing configurations. *Discrete Mathematics 204*, 1-3 (1999), 203–229. (Selected papers in honor of Henry W. Gould).

[41] FLAJOLET, P., AND ODLYZKO, A. The average height of binary trees and other simple trees. *Journal of Computer and System Sciences 25* (1982), 171–213.

[42] FLAJOLET, P., AND ODLYZKO, A. Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations. *Mathematical Proceedings of the Cambridge Philosophical Society 96* (1984), 237–253.

[43] FROBENIUS, G. Über Matrizen aus nicht negativen Elementen. *Sitz.-Ber. Akad. Wiss., Phys-Math Klasse, Berlin* (1912), 456–477.

[44] FULTON, W. *Algebraic Curves*. W.A. Benjamin, Inc., New York, Amsterdam, 1969.

[45] GANTMACHER, F. R. *Matrizentheorie*. Deutscher Verlag der Wissenschaften, Berlin, 1986. A translation of the Russian original Teoria Matriz, Nauka, Moscow, 1966.

[46] GODSIL, C. D. *Algebraic Combinatorics*. Chapman and Hall, 1993.

[47] GOOD, I. J. Random motion and analytic continued fractions. *Proceedings of the Cambridge Philosophical Society 54* (1958), 43–47.

[48] GOULDEN, I. P., AND JACKSON, D. M. *Combinatorial Enumeration*. John Wiley, New York, 1983.

[49] GOURDON, X. Largest component in random combinatorial structures. *Discrete Mathematics 180*, 1-3 (1998), 185–209.

[50] GRAHAM, R., KNUTH, D., AND PATASHNIK, O. *Concrete Mathematics*. Addison Wesley, 1989.

[51] HENRICI, P. *Applied and Computational Complex Analysis*, vol. 2. John Wiley, New York, 1974.

[52] HILLE, E. *Analytic Function Theory*. Blaisdell Publishing Company, Waltham, 1962. 2 Volumes.

[53] JACKSON, D. Some results on product-weighted lead codes. *Journal of Combinatorial Theory,* Series A *25* (1978), 181–187.

[54] KARLIN, S., AND TAYLOR, H. *A First Course in Stochastic Processes*, second ed. Academic Press, 1975.

[55] KIRSCHENHOFER, P., AND PRODINGER, H. The number of winners in a discrete geometrically distributed sample. *The Annals of Applied Probability 6*, 2 (1996), 687–694.

[56] KIRWAN, F. *Complex Algebraic Curves*. No. 23 in London Mathematical Society Student Texts. Cambridge University Press, 1992.

[57] KNOPFMACHER, A., AND PRODINGER, H. On Carlitz compositions. *European Journal of Combinatorics 19*, 5 (1998), 579–589.

[58] KNUTH, D. E. *The Art of Computer Programming*, vol. 1: Fundamental Algorithms. Addison-Wesley, 1968. Second edition, 1973.

[59] KNUTH, D. E. *The Art of Computer Programming*, 3rd ed., vol. 1: Fundamental Algorithms. Addison-Wesley, 1997.

[60] KRASNOSELSKII, M. *Positive solutions of operator equations*. P. Noordhoff, Groningen, 1964.

[61] LABELLE, J., AND YEH, Y.-N. Generalized Dyck paths. *Discrete Mathematics 82* (1990), 1–6.

[62] LAGARIAS, J. C., ODLYZKO, A. M., AND ZAGIER, D. B. On the capacity of disjointly shared networks. *Computer Netwoks 10*, 5 (1985), 275–285.

[63] LALLEY, S. P. Finite range random walk on free groups and homogeneous trees. *Ann. Probab. 21*, 4 (1993), 2087–2130.

[64] LANG, S. *Linear Algebra*. Addison-Wesley Publishing Co., Inc., Reading, Mass.-Don Mills, Ont., 1966.

[65] LENARD, A. Exact statistical mechanics of a one–dimensional system with Coulomb forces. *The Journal of Mathematical Physics 2*, 5 (Sept. 1961), 682–693.

[66] LOTHAIRE, M. *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and its Applications*. Addison–Wesley, 1983.

[67] LOUCHARD, G. Random walks, Gaussian processes and list structures. *Theoretical Computer Science 53*, 1 (1987), 99–124.

[68] MACMAHON, P. A. *Introduction to combinatory analysis*. Chelsea Publishing Co., New York, 1955. A reprint of the first edition, Cambridge, 1920.

[69] MEIR, A., AND MOON, J. W. On the altitude of nodes in random trees. *Canadian Journal of Mathematics 30* (1978), 997–1015.

[70] MEIR, A., AND MOON, J. W. The asymptotic behaviour of coefficients of powers of certain generating functions. *European Journal of Combinatorics 11* (1990), 581–587.

[71] MOHANTY, S. G. *Lattice path counting and applications*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1979. Probability and Mathematical Statistics.

[72] ODLYZKO, A. M. Periodic oscillations of coefficients of power series that satisfy functional equations. *Advances in Mathematics 44* (1982), 180–205.

[73] ODLYZKO, A. M. Asymptotic enumeration methods. In *Handbook of Combinatorics*, R. Graham, M. Grötschel, and L. Lovász, Eds., vol. II. Elsevier, Amsterdam, 1995, pp. 1063–1229.

[74] ODLYZKO, A. M., AND WILF, H. S. The editor's corner: n coins in a fountain. *American Matematical Monthly 95* (1988), 840–843.

[75] OTTER, R. The number of trees. *Annals of Mathematics 49*, 3 (1948), 583–599.

[76] PERCUS, J. K. *Combinatorial Methods*, vol. 4 of *Applied Mathematical Sciences*. Springer-Verlag, 1971.

[77] PERRON, O. Über Matrizen. *Mathematische Annalen 64* (1907), 248–263.

[78] PÓLYA, G. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathematica 68* (1937), 145–254.

[79] PÓLYA, G., AND READ, R. C. *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*. Springer Verlag, New York, 1987.

[80] READ, R. C. The chord intersection problem. *Annals of the New York Academy of Sciences 319* (May 1979), 444–454.

[81] ROBERT, P. *Réseaux et files d'attente: méthodes probabilistes*, vol. 35 of *Mathématiques & Applications*. Springer, Paris, 2000.

[82] SALVY, B., AND ZIMMERMANN, P. GFUN: a Maple package for the manipulation of generating and holonomic functions in one variable. *ACM Trans. Math. Softw. 20*, 2 (1994), 163–167.

[83] SCHAEFFER, G. *Conjugaison d'arbres et cartes combinatoires aléatoires*. Ph.d. thesis, Université de Bordeaux I, Dec. 1998.

[84] SEDGEWICK, R., AND FLAJOLET, P. *An Introduction to the Analysis of Algorithms*. Addison-Wesley Publishing Company, 1996.

[85] SLOANE, N. J. A. *The On-Line Encyclopedia of Integer Sequences*. 2000. Published electronically at `http://www.research.att.com/ñjas/sequences/`.

[86] SLOANE, N. J. A., AND PLOUFFE, S. *The Encyclopedia of Integer Sequences*. Academic Press, 1995. Available electronically at `http://www.research.att.com/ñjas/sequences/`.

[87] STANLEY, R. P. *Enumerative Combinatorics*, vol. I. Wadsworth & Brooks/Cole, 1986.

[88] STANLEY, R. P. *Enumerative Combinatorics*, vol. II. Wadsworth & Brooks/Cole, 1998.

[89] STEIN, P. R., AND WATERMAN, M. S. On some new sequences generalizing the Catalan and Motzkin numbers. *Discrete Mathematics 26*, 3 (1979), 261–272.

[90] SZEGŐ, G. *Orthogonal Polynomials*, vol. XXIII of *Americam Mathematical Society Colloquium Publications*. A.M.S, Providence, 1989.

[91] TEMPERLEY, H. N. V. *Graph theory and applications*. Ellis Horwood Ltd., Chichester, 1981. Ellis Horwood Series in Mathematics and its Applications.

[92] TOUCHARD, J. Contribution a l'étude du problème des timbres poste. *Canadian Journal of Mathematics 2* (1950), 385–398.

[93] TUTTE, W. T. A census of planar maps. *Canadian Journal of Mathematics 15* (1963), 249–271.

[94] TUTTE, W. T. On the enumeration of planar maps. *Bull. Amer. Math. Soc. 74* (1968), 64–74.

[95] TUTTE, W. T. On the enumeration of four-colored maps. *SIAM Journal on Applied Mathematics 17* (1969), 454–460.

[96] TUTTE, W. T. Planar enumeration. In *Graph theory and combinatorics (Cambridge, 1983)*. Academic Press, London, 1984, pp. 315–319.

[97] VALLÉE, B. Dynamical sources in information theory: Fundamental intervals and word prefixes. *Algorithmica 29*, 1/2 (2001), 262–306.

[98] VON ZUR GATHEN, J., AND GERHARD, J. *Modern computer algebra*. Cambridge University Press, New York, 1999.

[99] WALL, H. S. *Analytic Theory of Continued Fractions*. Chelsea Publishing Company, 1948.

[100] WOODS, A. R. Coloring rules for finite trees, and probabilities of monadic second order sentences. *Random Structures Algorithms 10*, 4 (1997), 453–485.

[101] ZEILBERGER, D. Closed form (pun intended!). *Contemporary mathematics 143* (1988), 579–607.

[102] ZEILBERGER, D. Symbol-crunching with the transfer-matrix method in order to count skinny physical creatures. *Integers 0* (2000), Paper A9. Published electronically at `http://www.integers-ejcnt.org/vol0.html`.

Contents

Index