Analytic Analysis of Algorithms

Philippe Flajolet
INRIA Rocquencourt

September 16, 1992

[summary by Pierre Nicodème]

Abstract

Symbolic methods in combinatorial analysis permit to express directly the counting generating functions of wide classes of combinatorial structures. Asymptotic methods based on complex analysis permit to extract directly coefficients of structurally complicated generating functions without a need for explicit coefficient expansions.

Three major groups of problems relative to algebraic equations, differential equations, and iteration are presented. The range of applications includes formal languages, tree enumerations, comparison-based searching and sorting, digital structures, hashing and occupancy problems.

This summary is based on [2].

Introduction

Quicksort. The classical analysis of the Quicksort algorithm results in solving a recurrence based on the recursive structure of the algorithm,

$$\bar{Q}_n = p_n + \sum_{k=0}^{n-1} \pi_{n,k}[\bar{Q}_k + \bar{Q}_{n-1-k}]. \tag{1}$$

There $\bar{Q}_n$ is the expected number of comparisons, $\pi_{n,k}$ is the probability that the partitioning stage splits the file into two subfiles of sizes k and $n-1-k$, and the quantity p_n represents the cost for partitioning.

There is an alternative approach to this problem.

Introduce the *generating function* (GF) of the mean values

$$Q(z) = \sum_{n=0}^{\infty} \bar{Q}_n z^n, \tag{2}$$

and set similarly $p(z) = \sum_{n\geq 0} p_n z^n$. Then, the equation corresponding to recurrence (1) is

$$Q(z) = p(z) + 2\int_0^z Q(t)\,\frac{dt}{1-t}. \tag{3}$$

The solution of this equation is

$$Q(z) = \frac{1}{(1-z)^2}\int_0^z \frac{d}{dt}\{p(t)\}\,(1-t)^2 dt. \tag{4}$$

and with $p(z) = z^2/(1-z)^2$,

$$Q(z) = 2\frac{\log(1-z)^{-1}}{(1-z)^2} - \frac{2z}{(1-z)^2}. \tag{5}$$

If we expand $Q(z)$, we retrieve again the solution of recurrence (1) that involves the harmonic numbers.

The solution expressed by (5) can be used to produce direct asymptotic results from the generating function itself, without any need for explicit expansions. The key observation is that it suffices to examine the generating function *locally* near its *singularity* at $z = 1$ and apply systematic translation mechanisms.

The translation from the local singular behaviour of a function to the asymptotics of its coefficients is a powerful mechanism. General rules valid under simple conditions (analytic continuation) apply, like for instance, the relation

$$[z^n]\ \frac{1}{(1-z)^\alpha}(\log(1-z)^{-1})^k \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^k.$$

1. Symbolic Methods in Combinatorial Analysis

The very powerful symbolic methods in combinatorial analysis may be summarized as follows:

PRINCIPLE. A number of set-theoretic constructions like union, cartesian product, sequence set, cycle set, power set, substitution have direct translation into generating function equations. Thus, a counting problem which is expressible in the language of these constructions can be translated systematically (and automatically) into generating function equations.

Given a class $\mathcal{F}$ of combinatorial structures, we let $\mathcal{F}_n$ denote the collection of objects of size n, and set $F_n = \mathrm{card}(\mathcal{F}_n)$. The *ordinary generating function* (OGF) and *exponential generating function* (EGF) are defined respectively to be

$$F(z) = \sum_{n\geq 0} F_n z^n \qquad \text{and} \qquad \hat{F}(z) = \sum_{n\geq 0} F_n \frac{z^n}{n!}. \tag{6}$$

A combinatorial construction is *admissible* if it admits a translation into generating functions.

THEOREM 1 (ADMISSIBLE CONSTRUCTIONS FOR OGF'S). *For unlabelled structures, the constructions of union, cartesian product, sequence, cycle, set, multiset, substitution are admissible. The translations into ordinary generating functions are given by the following table*

Construction	*Translation (OGF)*
$\mathcal{F} = \mathcal{G} \cup \mathcal{H}$	$F(z) = G(z) + H(z)$
$\mathcal{F} = \mathcal{G} \times \mathcal{H}$	$F(z) = G(z) \cdot H(z)$
$\mathcal{F} = \mathit{sequence}(\mathcal{G}) = \mathcal{G}^*$	$F(z) = \frac{1}{1-G(z)}$
$\mathcal{F} = \mathit{set}(\mathcal{G})$	$F(z) = \exp(G(z) - \frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) - \cdots)$
$\mathcal{F} = \mathit{multiset}(\mathcal{G})$	$F(z) = \exp(G(z) + \frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots)$
$\mathcal{F} = \mathit{cycle}(\mathcal{G})$	$F(z) = \log(1-G(z))^{-1} + \cdots$
$\mathcal{F} = \mathcal{G}[\mathcal{H}]$	$F(z) = G(H(z))$

THEOREM 2 (ADMISSIBLE CONSTRUCTIONS FOR EGF'S). *For labelled structures, the constructions of union, partitional product, sequence, cycle, set, substitution are admissible. The translations into exponential generating functions are given by the following table*

Construction	*Translation (EGF)*
$\mathcal{F} = \mathcal{G} \cup \mathcal{H}$	$\hat{F}(z) = \hat{G}(z) + \hat{H}(z)$
$\mathcal{F} = \mathcal{G} * \mathcal{H}$	$\hat{F}(z) = \hat{G}(z) \cdot \hat{H}(z)$
$\mathcal{F} = \mathit{sequence}(\mathcal{G}) = \mathcal{G}^*$	$\hat{F}(z) = \frac{1}{1-\hat{G}(z)}$
$\mathcal{F} = \mathit{set}(\mathcal{G})$	$\hat{F}(z) = \exp(\hat{G}(z))$
$\mathcal{F} = \mathit{cycle}(\mathcal{G})$	$\hat{F}(z) = \log(1-\hat{G}(z))^{-1}$
$\mathcal{F} = \mathcal{G}[\mathcal{H}]$	$\hat{F}(z) = \hat{G}(\hat{H}(z))$

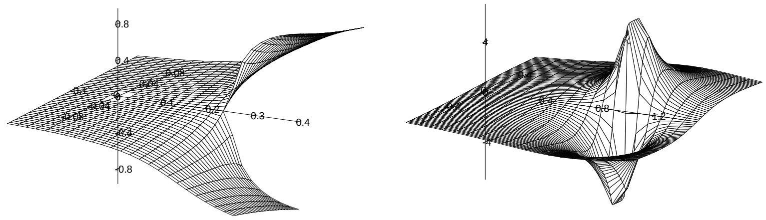

FIGURE 1
A display of the imaginary parts of two generating functions, $f(z) = \frac{1-\sqrt{1-4z}}{2z}$ and $g(z) = \frac{1}{1-z}$. The function $f(z)$ [left] is the ordinary generating function of binary trees with a singularity at $\rho = 1/4$ which is a branch point of the $\sqrt{\ }$ type. The function $g(z)$ [right] is the exponential generating function of permutations with a singularity at $\rho = 1$ of a polar type. The singularities are reflected at the level of coefficients, $[z^n]f(z) \sim \frac{4^n}{\sqrt{\pi n^3}}$ and $[z^n]g(z) = 1$.

2. Complex Analysis and Asymptotics

Complex analytic methods permit to represent coefficients of generating functions and many combinatorial sums as integrals of an analytic function in the complex plane. The choice of a suitable contour of integration often leads to highly non trivial asymptotic results.

Singularity analysis. Most functions occurring in combinatorial enumeration problems are built by operators from standard functions that exist over the whole of the complex plane. They thus tend to exist in larger areas of the complex plane. The method of *singularity analysis* is well suited to extracting coefficients of functions lying in a class that enjoys interesting closure properties.

Saddle point integrals. The saddle point method is useful for the computation of coefficients of whole classes of entire functions, with the following asymptotics:

THEOREM 3 (SADDLE POINT COEFFICIENT ASYMPTOTICS). *For a function $f(z)$ to which the saddle point method applies, one has*

$$[z^n]f(z) \sim \frac{f(\zeta)}{\sqrt{2\pi C}\zeta^{n+1}} \quad \textit{where} \quad C = \frac{d^2}{dz^2}\log\frac{f(z)}{z^{n+1}}\bigg|_{z=\zeta},$$

and $\zeta = \zeta_n$ is the smallest real root of $\frac{d}{dz}\log\frac{f(z)}{z^{n+1}}$.

Although leading to difficult questions, the method may be applied to dimensions higher than 1 dimensional saddle [3] [6].

3. Algebraic Functions and Implicit Functions

Regular languages can be specified either by regular expressions or by finite automata. The corresponding GF's either appear as built from the variable z by means of rational operations ($+$, $\times$, quasi-inverse $Q(y) = (1-y)^{-1}$) or as components of linear systems of equations (over $\mathbb{Z}[x]$). At any rate, they are rational.

An immediate consequence of the partial fraction decomposition of rational functions is the following.

THEOREM 4 (RATIONAL ASYMPTOTICS). *The coefficients of a rational function of $\mathbb{Q}(z)$ are a finite linear combination of 'exponential polynomials' of the form*

$$\lambda\,\omega^n n^k, \tag{7}$$

with λ, ω algebraic numbers and k an integer.

Context free languages lead to polynomial nonlinear equations, provided the grammar is unambiguous or we count words with their multiplicities. Thus, the generating function of a context free language is algebraic and the following theorem holds:

THEOREM 5 (ALGEBRAIC ASYMPTOTICS). *The coefficients of a $\mathbb{Q}(z)$-algebraic function are asymptotic to a sum of 'algebraic elements' of the form*

$$\frac{\lambda}{\Gamma(r/s+1)}\omega^n n^{r/s}, \tag{8}$$

where λ, ω are algebraic numbers, and the exponent r/s is a rational number.

Implicit functions. Functions defined implicitly tend to have singularities like those of algebraic functions, involving fractional exponents. This is reflected by the asymptotics of their coefficients of the form $\omega^n n^{-r/s}$. Such a property also holds for many functions satisfying finite and infinite functional equations involving terms like $f(z^2), f(z^3)$ provided that their radius of convergence is < 1.

4. Holonomic Functions and Differential Equations

Functions satisfying differential equations with polynomial coefficients are sometimes called $\mathcal{D}$-finite and their coefficient sequences which satisfy recurrences with polynomial (in n) coefficients are then called $\mathcal{P}$-recursive. These notions are formalized by the concept of *holonomy* introduced in this range of problems by Zeilberger.

DEFINITION 1. A series $f(z_1, z_2, \ldots, z_r) \in \mathbb{C}[[z_1, z_2, \ldots, z_r]]$ is said to be *holonomic* iff the infinite collection of its partial derivatives

$$\frac{\partial^{j_1}}{\partial z^{j_1}}\frac{\partial^{j_2}}{\partial z^{j_2}}\cdots\frac{\partial^{j_r}}{\partial z^{j_r}} f(z_1, z_2, \ldots, z_r)$$

spans a finite dimensional vector space over the field of rational fractions $\mathbb{C}(z_1, z_2, \ldots, z_r)$.

A sequence $f_{n_1, n_2, \ldots, n_r}$ is holonomic iff its generating function

$$f(z_1, z_2, \ldots, z_r) = \sum_{n_1, n_2, \ldots, n_r} f_{n_1, n_2, \ldots, n_r} z_1^{n_1} z_2^{n_2} \cdots z_r^{n_r}$$

is holonomic.

The major closure theorem here is due to Stanley, Lipschitz, and Zeilberger [4, 5, 7, 8].

THEOREM 6 (HOLONOMIC CLOSURE). *Holonomic functions are closed under sums, Cauchy products, Hadamard products, diagonals, algebraic substitutions, integration, differentiation, direct and inverse Laplace transforms.*

THEOREM 7 (HOLONOMIC ASYMPTOTICS). *A holonomic sequence f_n is asymptotic to a sum of elements of the form*

$$\lambda(n!)^{r/s} e^{Q(n^{1/m})} \omega^n n^\alpha (\log n)^k,$$

where r, s, m, k are integers, Q is a polynomial and λ, ω, α are complex numbers.

In our perspective, this theorem relates to the classification of singularities of linear differential equations. The theory of linear differential equations with analytic coefficients distinguishes for solutions of such equations two cases, the regular case and the irregular case. The method of singularity analysis and the method of saddle point integrals are applicable each in one of the two cases.

For instance the expected cost of a partial match query in a quadtree (alternatively a *k-d*-tree) when a proportion of $\frac{1}{2}$ or $\frac{2}{3}$ of the coordinates is known is of the order of

$$n^{(\sqrt{17}-3)/2} \qquad \text{and} \qquad n^{\theta-1} \quad \text{with} \quad \theta = \left(\frac{109}{27} + \sqrt{\frac{1320}{81}}\right)^{1/3} + \left(\frac{109}{27} - \sqrt{\frac{1320}{81}}\right)^{1/3}.$$

Such algebraic numbers in the exponents are typical of $\mathbb{Q}(z)$ holonomic functions.

5. Functional Equations and Iteration

We confine our discussion to *linear functional equations* of the form

$$f(z) = a(z) + b(z)f(\sigma(z)), \tag{9}$$

where $f(z)$ is the unknown function, and a, b, σ are explicitly known. In the functional equation of (9), everything depends crucially on the dynamics of the iterates of σ. In a few important cases, the iterates are explicit, and one general method available relies on the Mellin transform.

Explicit iterations. The analysis of digital tries furnishes an example of the situation where the iteration of $\sigma(z)$ is explicit. The recurrence of expected path length in tries is of a probabilistic divide-and-conquer type,

$$f_n = n - \delta_{n,1} + 2\sum_{k=0}^{n} \pi_{n,k} f_k \qquad \text{with } \pi_{n,k} = \frac{1}{2^n}\binom{n}{k}.$$

The corresponding EGF satisfies

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

The equation is solved by iteration, after which the solution can be expanded.

The method is also applicable to wide classes of divide-and-conquer recurrences which are almost invariably found to give rise to periodic fluctuations involving fractals.

Implicit iterations. When the iterates $\sigma^{\langle j\rangle}(z)$ admit of no simple explicit form, one often has to resort to an analysis of individual terms in the sum (9), normally by the battery of complex analysis techniques examined so far.

At the moment, a complete classification of the various cases of (9) is still lacking. Some cases appear to involve the theory of analytic iteration and some divergent series. We nonetheless have a number of useful and general tools available in the form of Mellin transforms and iteration theory of analytic functions.

6. Automatic Analysis

The approach of finding general decidable asymptotic properties of combinatorial structures has been prolonged. Flajolet, Salvy and Zimmermann [1] have designed a system called Lambda-Upsilon-Omega ($\Lambda\Upsilon\Omega$) that implements a number of decision procedures on combinatorial structures like the ones discussed here. The kernel specification language consists of the constructions of union, product, sequence, sets, multisets and cycles described in Section 1. The $\Lambda\Upsilon\Omega$ system also makes provisions for specifying traversal algorithms on the structures.

Bibliography

[1] Flajolet (P.), Salvy (B.), and Zimmermann (P.). – Automatic average-case analysis of algorithms. *Theoretical Computer Science, Series A*, vol. 79, n° 1, February 1991, pp. 37–109.
[2] Flajolet (Philippe). – Analytic analysis of algorithms. In Kuich (W.) (editor), *Automata, Languages and Programming*, *Lecture Notes in Computer Science*, pp. 186–210. – 1992. Proceedings of the 19th International Colloquium, Vienna, July 1992. (Invited lecture).
[3] Gardy (Danièle). – Méthode de col et lois limites en analyse combinatoire. *Theoretical Computer Science*, vol. 92, n° 2, 1992, pp. 261–280.
[4] Lipshitz (L.). – The diagonal of a D-finite power series is D-finite. *Journal of Algebra*, vol. 113, 1988, pp. 373–378.
[5] Lipshitz (L.). – D-finite power series. *Journal of Algebra*, vol. 122, 1989, pp. 353–373.
[6] McKay (Brendan D.). – The asymptotic numbers of regular tournaments, Eulerian digraphs and Eulerian oriented graphs. *Combinatorica*, vol. 10, n° 4, 1990, pp. 367–377.
[7] Stanley (R. P.). – Differentiably finite power series. *European Journal of Combinatorics*, vol. 1, 1980, pp. 175–188.
[8] Zeilberger (Doron). – A holonomic systems approach to special functions identities. *Journal of Computational and Applied Mathematics*, vol. 32, 1990, pp. 321–368.