INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Algorithms Seminar, 2002–2004

Frédéric CHYZAK, éditeur scientifique

N ° 5542
Avril 2005

THÈME 2

ISSN 0249-6399

Algorithms Seminar, 2002–2004

Frédéric CHYZAK, éditeur scientifique

Thème 2 — Génie logiciel
et calcul symbolique
Projet Algo

Rapport de recherche n ° 5542 — Avril 2005 — 116 pages

Abstract: These seminar notes constitute the proceedings of a seminar devoted to the analysis of algorithms and related topics. The subjects covered include combinatorics, symbolic computation, and the asymptotic analysis of algorithms, data structures, and network protocols.

Key-words: combinatorics, symbolic computation, analysis of algorithms

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

Séminaire algorithmes, 2002–2004

Résumé : Ces notes de séminaires constituent les actes, le plus souvent en anglais, d'un séminaire consacré à l'analyse d'algorithmes et à ses domaines connexes. Les thèmes abordés comprennent la combinatoire, le calcul formel, l'analyse asymptotique d'algorithmes, de structures de données et de protocoles de réseaux.

Mots-clé : combinatoire, calcul formel, analyse d'algorithmes

ALGORITHMS SEMINAR 2002–2004

Frédéric Chyzak
(Editor)

These seminar notes constitute the proceedings of a seminar whose primary goal is to cover the major methods for the average-case analysis of algorithms and data structures. Neighbouring topics of study are combinatorics, symbolic computation, network protocols, asymptotic analysis, probabilistic methods, and dynamical systems. The contents of these recurrent proceedings consist of summaries of the talks, written by members of the audience.[1]

The study of combinatorial objects—their description, their enumeration according to various parameters—arises naturally in the process of analysing algorithms that often involve classical combinatorial structures like strings, permutations, partitions, trees, graphs, and maps.

Asymptotic analysis is an essential ingredient in the interpretation of quantitative results supplied by the resolution of combinatorial models. Various asymptotic methods are found to be relevant to the analysis of particular algorithms. These proceedings include singularity analysis, the saddle-point method in presence of coalescence, functional iteration, and an approach by dynamical systems.

As already said, the average-case asymptotic analysis of algorithms is the primary goal of the seminar. Interestingly enough, the study of the combinatorics or probabilistic process that underly an algorithm sometimes leads to algorithmic improvements. These proceedings show analyses of state-of-the-art algorithms and report on complexity-driven designs of algorithms and protocols.

Symbolic computation, and in particular computer algebra, plays an increasingly important role in these areas. It provides a collection of tools that allow one to attack complex models of combinatorics and the analysis of algorithms via generating functions; at the same time, it inspires the quest for developing ever more systematic solutions and decision procedures for the analysis of well-characterized classes of problems.

In the past, it had been customary to publish one set of proceedings a year, collecting summaries for almost all of the talks given at the seminar in the entervening period of time. This rythm had indeed been maintained since 1991 for a period of 11 years. Yet, it appeared that volunteers could recently less easily find the time to write their summaries. The 18 articles and one set of Aléa'04 course notes included in this book represent only a rough third of the real activity of the seminar in the three-year period 2002–2004, and a specially low proportion of talks on computer algebra algorithms. They are nevertheless snapshots of very recent research in the areas mentioned above. A tentative organization of the contents is given below.

PART I. ENUMERATIVE COMBINATORICS

Combinatorial maps have received an important treatment at the seminar over the period of time. The enumeration bipartite planar maps according to their vertex degrees is the topic of [**1**] and that of unrooted maps on the sphere the topic of [**6**]. In relation to this, we have included notes for a course given at the Aléa'04 workshop, [**7**]. There, two approaches to the enumeration of planar maps are described: by bijective combinatorics and by generating functions. Knot

[1] The summaries for the past years are also available on the web at `http://algo.inria.fr/seminars/`.

theory and Feynman diagrams also lead to models of maps. The underlying enumerations remain untractable except for special cases, but conjectures can be tested by generating random maps [**4**]. Other physical models are analysed in [**3**] by introducing a model of planar triangulations. A generalization of partitions, called overpartitions, and a constrained model of pairs of overpartitions, called particle seas, are introduced in [**5**] to explain q-summations and products of the theory of partitions by bijective means. Many combinatorial generating functions that encode models with symmetries can be expressed as combinations of symmetric functions or extracted as sub-series from such combinations. A symbolic algorithm for such extractions is the topic of [**2**].

[**1**] The Degree Distribution of Bipartite Planar Maps and the Ising Model. *G. Schaeffer.*
[**2**] Effective Scalar Product of Differentiably Finite Symmetric Functions. *F. Chyzak.*
[**3**] Heaps of Segments and Lorentzian Quantum Gravity. *W. James.*
[**4**] Matrix Models an Knot Theory. *P. Zinn-Justin.*
[**5**] Particle Seas and Basic Hypergeometric Series. *S. Corteel.*
[**6**] Counting Unrooted Maps Using Tree Decomposition. *É. Fusy.*
[**7**] Deux Approches pour l'Énumération des Cartes planaires *(Two Approaches to the Enumeration of Planar Maps). G. Schaeffer.*

PART II. ANALYTIC COMBINATORICS AND ASYMPTOTICS

Urn models of the Pólya type describe evolutions of several kinds of coexisting populations. An analytic approach to determine the asymptotic behaviour of such urns when the total population grows uniformly and linearly in time is described in [**10**]. The topic is continued in [**11**], where the involved PDEs are solved more explicitly and more precisely. Random recursive trees are a classical model for the growth of trees. The asymptotic profile of such trees is analysed in [**14**], where consequences on the asymptotic profile of random binary search trees are also drawn. Suffix trees and tries are data structures well-suited for the implementation and analysis of dictionaries, compression algorithms, and certain communication protocols. A comparison of the average size of suffix trees and tries in the non-uniform Bernoulli model is performed in [**12**]. It is quite a natural idea to consider an algorithm together with its possible inputs as a dynamical system. In the analysis of algorithms and dynamical systems by this recent methodology, constants which arise can usually be determined only numerically. A polynomial-time algorithm is described in [**13**] for the computation of guaranteed truncations of such constants. The saddle-point method is used in presence of two coalescing saddle points in [**15**] to give a new treatment of the enumeration of connected graphs by excess (number of edges minus number of vertices). It is shown in [**9**] that the number of occurrences of a given pattern in general random trees is asymptotically normal with the size of the searched tree. This result is obtained by a nice description of the autocorrelation of the pattern tree, leading to a system of differential equations amenable to a general asymptotic theorem of Drmota's. An asymptotic behaviour related to a fixed-point combinatorial equation involving a class subtitution is analysed in [**8**], basing on analytic iteration theory.

[**8**] On the Asymptotic Analysis of a Class of Linear Recurrences. *T. Prellberg.*
[**9**] Patterns in Trees. *T. Klausner.*
[**10**] Analytic Urns. *Ph. Flajolet.*
[**11**] Analytic Urns of Triangular Form. *V. Puyhaubert.*
[**12**] Suffix Trees and Simple Sources. *J. Fayolle.*
[**13**] Efficient Computation of a Class of Continued Fraction Constants. *L. Lhote.*
[**14**] Profile of Random Recursive Trees and Random Binary Search Trees. *H.-K. Hwang.*
[**15**] Airy Phenomena and the Number of Sparsely Connected Graphs. *B. Salvy.*

PART III. ANALYSIS OF ALGORITHMS AND PROTOCOLS

Estimating the number of distinct words which appear repeated in a massive flow of data has important practical applications, e.g., for measuring the activity of a router on a network. A new probabilistic algorithm that uses little memory and few processor cycles for each piece of data is analysed in [**17**]. 'Quicksort' and 'quickselect' are classical divide-and-conquer algorithms to respectively sort an array and to extract its m-th smallest entry. The constants in their complexity estimates depends on the choice of a good pivot. Various strategies are surveyed in [**19**] and a new efficient algorithm is introduced there to select and sort the m smallest entries of an array. The election of a leader is a fundamental problem in distributed systems. For instance, consider an ad-hoc network in which a subset of the peers needs to be selected to provide specialized services for the remaining peers. An efficient and economical probabilistic algorithm for this task is given in [**18**]. The well-known algorithm by Knuth to draw permutations uniformly at random has been adapted by Sattolo to draw cyclic permutations. Two probabilistic approaches to its asymptotic analysis are summarized in [**20**], together with a new approach by methods of statistical physics. Gröbner bases are the most fundamental tool in (multivariate) computational commutative algebra. Notwithstanding the recent algorithmic improvements for their calculations, leading to tremendous practical speed-ups, little was known on the theoretical side about the generic behaviour of the corresponding algorithms. The current best algorithm in practice is analysed in [**16**], showing a generic behaviour far below the worst-case EXPSPACE complexity. The analysis focusses on two coalescing saddle points.

Acknowledgements. The lectures summarized here emanate from a seminar attended by a community of researchers from the Algorithms Project at INRIA and the greater Paris area. Its organization is conducted by Frédéric Chyzak, Philippe Flajolet, and Bruno Salvy. The editor expresses his gratitude to the various people who have actively supported this joint enterprise. Thanks are due to the speakers and to the authors of summaries. Again in this edition, roughly half of the texts are summaries by researchers from the Algorithms Project. We are also greatly indebted to our secretary Virginie Collette for making all of the organization work smoothly.

The editor,
F. CHYZAK

Part I

Enumerative Combinatorics

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 3–6.

Available online at the URL
http://algo.inria.fr/seminars/.

The Degree Distribution of Bipartite Planar Maps and the Ising Model

Gilles Schaeffer

Lix, École Polytechnique (France)

March 10, 2003

Summary by Julien Fayolle

Abstract

Enumerating bipartite (with black and white vertices) planar maps according to the degree distribution of the vertices is useful to physicists. We first exhibit a bijection between these maps and some family of trees. The generating functions of these trees are then obtained with classical decomposition on the combinatorial structure of the trees.

The physicists need to add *Ising* or *hard particle* models to planar maps to model particle location or spin. We can relate bijectively these maps with an additional structure to the bipartite maps. We finally enumerate the Ising and hard particle configurations on maps. (Joint work with Mireille Bousquet-Mélou from Labri)

1. Introduction

Following their bijective instincts the authors use in [1] a combinatorial approach to solve problems arising from 2-dimensional *quantum gravity*: the enumeration of planar maps under *Ising* and *hard particle* models. The main results have already been obtained by use of the powerful matrix integrals approach [3, 2]. Alas the matrix integral does not provide any insight on the combinatorial behavior of these maps like why their generating functions are algebraic. The authors exhibit a bijection between bipartite planar maps and some class of trees to explain the algebraicity of the generating functions. They then show the relation between bipartite planar maps and maps with the physicists' additional models.

Let's first recall some basic definitions: a *planar map* is a connected graph drawn on the sphere with non-intersecting edges, a *rooted map* is a map that possesses a root edge, i.e., an oriented edge. As a convention, when the map is drawn on the plane, its infinite face lies to the right of the rooted edge. In this talk we only deal with bipartite maps, meaning that each and every edge of the map has one black and one white endpoint.

The *degree distribution* of a bipartite map is a pair of partitions (λ, μ) coding the number of white vertices of a certain degree and the number of black vertices of a certain degree. The ith part of the partition λ is the number of white vertices of degree i. For example the map on Figure 2 has a contribution $x_2^2 x_3 x_4^2 y_2^2 y_4^2 y_5$ to the generating function where x_i counts the number of white vertices of degree i and y_i the number of black vertices of degree i.

2. Maps and Trees

Direct enumeration of bipartite planar maps being difficult, we try a bijective argument: since planar maps are closely related to trees we enumerate a certain class of tree. Let's first introduce *blossom* trees: these are trees with black-and-white edges and also *half-edges*, which are called *leaves*

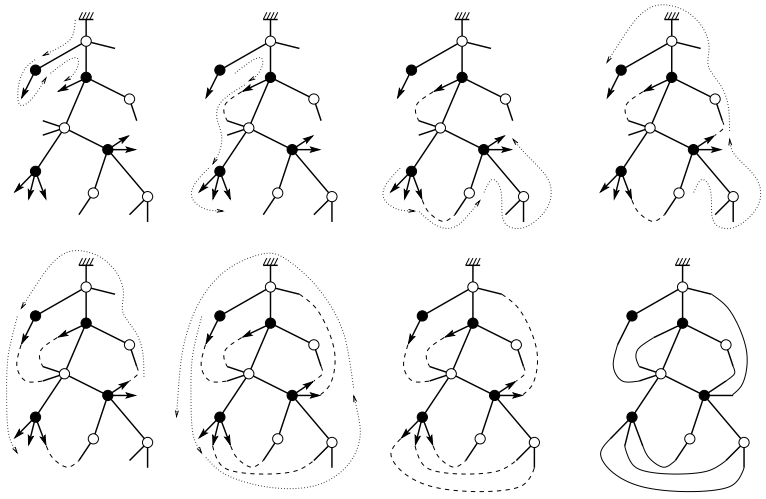


FIGURE 1. Closure

or *buds* whether they hang from a white or a black vertex. These trees are rooted at a leaf or a bud, and the vertex attached to this half-edge is called the root vertex. Furthermore, we define the *total charge* (resp. *charge*) of a blossom tree as the difference between the number of leaves in the tree and the number of buds where the root half-edge does (resp. does not) count. The charge of a vertex is just the charge of the subtree rooted at this vertex. A blossom tree must furthermore satisfy the condition that all white vertices have nonnegative (≥ 0) charge and all black ones a charge at most 1 (except the root vertex which has no charge condition to satisfy). A *k-leg map* is a bipartite map rooted at a half-edge with k leaves and no bud (hence the map is rooted at a leaf).

2.1. **Closure.** We define a closure operation ϕ on blossom trees T of total charge $k \geq 1$ by walking counterclockwise around the infinite face. Each and every time the next half-edge after a bud is a leaf we fuse these two half-edges into an edge. This process eventually stops and we obtain a bipartite planar map rooted at a half-edge with k leaves that remain unmatched, i.e., a k-leg map. Last but not least we introduce *balanced* trees: these trees are rooted at a leaf and after closure, the root is unmatched.

Proposition 1. *Let T be a balanced tree having total charge k ($k \geq 1$) and degree distribution (λ, μ). Then $\phi(T)$ is a k-leg map with degree distribution (λ, μ).*

Half-edges are counted as regular edges when determing the degree of a vertex. This is highly useful here since it keeps the degree distribution of the vertices unchanged by fusing half-edges (and also by splitting a bicolor edge as we will see in the next paragraph).

2.2. **Opening.** We shall now define the reciprocal operation of the mapping ϕ: the opening ψ of a k-leg map. Let M be a k-leg map, we walk counterclockwise around the infinite face starting from the root. For each edge visited from its black to its white endpoint that can be deleted without breaking the connectedness of the map, we cut this edge into two half-edges: a bud from the black endpoint and a leaf from the white one. At the end of the process we get a balanced tree (since it's rooted at a leaf (root of the k-leg map) and its root is a single leaf) of *total charge* (for this kind of charge we also count the rooted leaf) k.

Theorem 1. *Closure and opening are inverse bijection between balanced blossom trees of total charge k and k-leg maps. Moreover the degree distribution is preserved.*

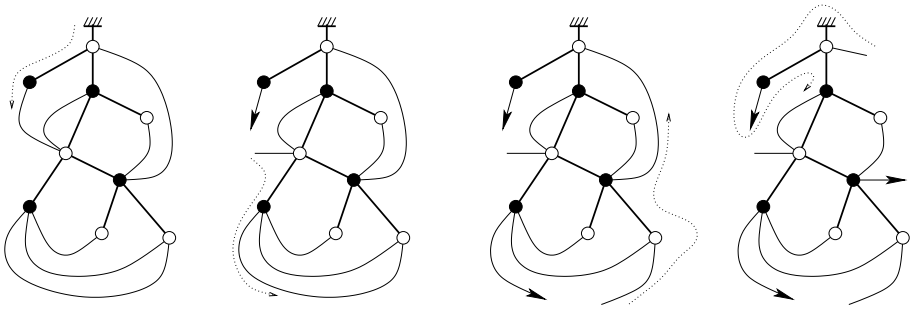


FIGURE 2. Opening

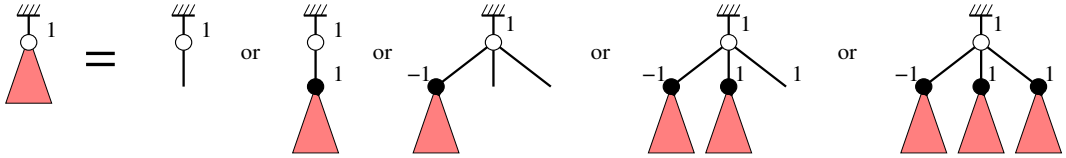


FIGURE 3. Decomposition

The reader is referred to [1] for a complete proof. We know that the generating functions of trees are algebraic, and this holds for balanced blossom trees of charge k. The bijection between blossom trees of total charge k and bipartite k-leg maps assures us that the generating function of k-leg maps is also algebraic.

3. Counting some Blossom Trees

We enumerate blossom trees according to the charge and the color of their root vertex. The methods are overall the same and we focus here on the degree generating function $A_1^{\circ-}(\mathbf{x},\mathbf{y})$ of blossom trees of charge 1 rooted at a white vertex (hence the "$\circ-$" exponent) where furthermore the vertices can have degree 2 or 4. We have a decomposition of these trees involving other blossom trees with degrees 2 or 4 as can be seen on Figure 3.

If the root vertex has degree 2 then one edge is the root leaf and the other one is either a leaf (charge 1) or a black and white edge (remember the tree is bicolor) with the charge of the tree hanging from the black endpoint being 1. If the root vertex has degree four, things follow the same pattern. We obtain a system with four degree generating functions:

$$A_1^{\circ-} = x_2(1+A_1^{\bullet-}) + 3x_4A_{-1}^{\bullet-}(1+A_1^{\bullet-})^2, \qquad A_3^{\circ-} = x_4(1+A_1^{\bullet-})^3,$$
$$A_1^{\bullet-} = y_2A_1^{\circ-} + 3y_4(A_3^{\circ-} + (A_1^{\circ-})^2), \qquad A_{-1}^{\bullet-} = y_2 + 3y_4A_1^{\circ-}.$$

We solve this system to get $A_1^{\circ-}(\mathbf{x},\mathbf{y})$.

4. Ising and Hard Particle Models

We add an Ising configuration to a graph $\mathcal{G}$ by coloring its vertices with two colors (black and white). An edge is said *frustrated* if it has end points of the two colors. The hard particle model is a particularization of the Ising model where we forbid edges with two black endpoints ($\bullet-\bullet$). This means that no two physical particle can be adjacent in the graph.

We look for a bijection between maps with these two additional physical models (on the edges) and the bipartite maps we already obtained results on. On bipartite maps both $\circ-\circ$ and $\bullet-\bullet$

edges are forbidden. We use a trick to go from maps with a hard particle configuration to bipartite maps: we transform the $\circ - \circ$ edges of the map with a hard particle configuration into $\circ - \spadesuit - \circ$ edges in the bipartite map. The spades vertices are counted as ordinary black edges in the bipartite map.

The generating function related to the physics of the hard particle model is

$$H(\mathbf{X}, \mathbf{Y}, u) = \sum_{\mathcal{G}} \mathbf{X}^{\circ(\mathcal{G})} \mathbf{Y}^{\bullet(\mathcal{G})} u^{\circ - \circ(\mathcal{G})},$$

where each exponent respectively counts the degree of white vertices, black vertices and the number of $\circ - \circ$ edges in the map. The $\circ - \circ$ edges are said *vacant* because there is no physical particle on any end of the edge.

We want to obtain the generating function $H(\mathbf{X}, \mathbf{Y}, u)$ of maps with a hard particle configuration when the map is rooted at a vacant edge. Therefore when we use the transformation trick the bipartite map must be rooted at a black vertex of degree 2. We know from the previous sections how to obtain the degree generating function $F(\mathbf{x}, \mathbf{y})$ of bipartite maps rooted at a black vertex of degree 2.

Theorem 2. *The degree generating function $F(\mathbf{x}, \mathbf{y})$ of bipartite planar maps rooted at a black vertex of degree 2 is related to generating functions of blossom trees:*

$$F(\mathbf{x}, \mathbf{y}) = y_2((A_0^{\circ -} - A_2^{\bullet -})^2 + A_1^{\circ -} - A_3^{\bullet -} - (A_2^{\bullet -})^2).$$

Once we have this degree generating function [note that the generating functions $A_i^{\circ -}$ and $A_i^{\bullet -}$ are not the same as in section 3 since we have no condition on the degree of the vertices here], we notice that the transformation trick does not modify the degree of the white vertices hence $X_k = x_k$. The black vertices of degree $k \neq 2$ keep the same degree and the same color (these vertices are of the $\bullet$ type, not of the $\spadesuit$ type) thus $Y_k = y_k$. The black vertices of degree 2 can be of the two types: either it is a $\spadesuit$ vertex and then reversing the trick will transform it into a vacant edge marked by the variable u, or it is a $\bullet$ vertex and the degree remains the same. The leading y_2 in $F(\mathbf{x}, \mathbf{y})$ stems from the root condition and the black vertex at the root is of the $\spadesuit$ kind so we substitute $y_2 = u$ here. For the rest we substitute $y_2 = u + Y_2$ in the blossom trees generating functions. Once we applied the four substitutions in $F(\mathbf{x}, \mathbf{y})$ we obtain the degree generating function $H(\mathbf{X}, \mathbf{Y}, u)$.

For maps with the Ising model, we use a generalization of the transformation trick for the hard particle model, using two kind of white vertices. The substitution in the generating function counting bipartite maps is a little bit more complicated but the same method applies.

Bibliography

[1] Bousquet-Mélou (Mireille) and Schaeffer (Gilles). – The degree distribution in a bipartite planars maps: applications to the Ising model, 2002. `http://xxx.lpthe.jussieu.fr/abs/math.CO/0211070` also as an extended abstract in FPSAC'03.

[2] Di Francesco (Philippe), Ginsparg (Paul), and Zinn-Justin (Jean). – 2d gravity and random matrices. *Physics Reports*, vol. 254, n° 1–2, 1995.

[3] Zvonkine (Alexandr). – Matrix integrals and map enumeration: an accessible introduction. *Mathematical and Computer Modelling*, vol. 26, n° 8–10, 1997, pp. 281–304.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 7–10.

Available online at the URL
http://algo.inria.fr/seminars/.

Effective Scalar Product of Differentiably Finite Symmetric Functions

Frédéric Chyzak
Algorithms Project, INRIA (France)

March 11, 2003

Summary by Marni Mishna

Introduction

Many enumerative problems can be expressed using the scalar product of symmetric functions. As we shall see, we can set up expressions for the generating functions of objects which possess a certain kind of regularity as a scalar product. Two examples of this are k-regular graphs (graphs in which each vertex is of degree k), and secondly we have a class of semi-standard Young tableaux in which each entry appears k times. The generating series for these are respectively given by:

$$R_k(t) = \left\langle \left(\exp \sum_{i\geq 1} p_i/i\right)[e_2], \exp(h_k t) \right\rangle, \qquad Y_k(t) = \left\langle \left(\exp \sum_{i\geq 1} p_i/i\right)[e_1+e_2], \frac{1}{1-h_k t} \right\rangle.$$

Gessel [2] proved that this operation preserves some notion of D-finiteness, which implies that these generating functions should satisfy a linear differential equation with polynomial coefficients. This seminar, and [1] introduce algorithms to determine these differential equations, and prove that the algorithm is correct, and terminates.

1. Symmetric Functions

A *partition* of a positive integer n is a decreasing sequence of integers $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ whose sum is n. This is denoted $\lambda \vdash n$. A partition is either written as a vector or in power notation, for example $(7,7,4,4,1) = \langle 7^2 4^2 1\rangle$ are both partitions of 23. A *symmetric function* is a sum of monomials in a some variable set, such that it is invariant under any permutation of that variable set. We can write any symmetric function as a sum of *monomial symmetric functions*, defined for the variable set $\{x_1, x_2, \ldots\}$ with respect to some partition λ as

$$m_\lambda := \sum_{\sigma\in\mathcal{S}_{\mathbb{N}\setminus\{0\}}} (r_1!\, r_2! \cdots)^{-1}\, x_{\sigma(1)}^{\lambda_1} \cdots x_{\sigma(k)}^{\lambda_k}.$$

For example, $m_{(3,2,2)} = x_1^3x_2^2x_3^2 + x_3^3x_2^2x_1^2 + x_4^3x_1^2x_3^2 + \cdots$. We also have the *elementary symmetric functions*, $e_n = m_{\langle 1^n\rangle}$, and $e_\lambda = e_{\lambda_1}\cdots e_{\lambda_k}$; the homogeneous symmetric functions $h_n = \sum_{\lambda\vdash n} m_\lambda$, and $h_\lambda = h_{\lambda_1}\cdots h_{\lambda_k}$; and *power symmetric functions* $p_n = m_{(n)} = x_1^n + x_2^n + \ldots$, $p_\lambda = p_{\lambda_1}\cdots p_{\lambda_k}$. There is much to be said of the ring of symmetric functions, and the interested reader is pointed to Stanley [3, Chapter 7] for the intimate details. The essential here is that symmetric functions can always be written in terms of the power sum symmetric functions, and thus we have the liberty of describing our algorithms and definitions for symmetric functions as elements of $\Lambda = \mathbb{Q}[p_1, p_2 \ldots]$. In fact, we shall take this one step further and interest ourselves with the ring of series of symmetric functions, $\hat{\Lambda} = \mathbb{Q}[[p_1, p_2, \ldots]]$.

The scalar product of symmetric functions is the bilinear product defined by $\langle p_\lambda, p_\mu\rangle = z_\lambda \delta_{\lambda,\mu}$, with $\lambda = \langle 1^{r_1} 2^{r_2} \cdots\rangle$ giving $z_\lambda = (1^{r_1} r_1!)(2^{r_2} r_2!)\cdots$. It permits the extraction of the coefficients in symmetric function:

$$[x_1^{\lambda_1} x_2^{\lambda_2} \dots x^{\lambda_k}]\,\phi = [m_\lambda]\,\phi = \langle h_\lambda, \phi\rangle.$$

This is precisely the use we have in mind.

To conclude this all-too-brief introduction of symmetric functions, we describe one type of composition that turns out to be quite useful here: *plethysm*, written $f[g]$. We can most easily define it using the power sum symmetric functions. It is defined by $p_n[\psi(p_1, p_2, \dots)] = \psi(p_n, p_{2n}, \dots)$, along with $(\phi_1 + c\phi_2)[\psi] = \phi_1[\psi] + c\phi_2[\psi]$ and $(\phi_1 \cdot \phi_2)[\psi] = \phi_1[\psi] \cdot \phi_2[\psi]$.

2. D-Finite Functions and Holonomic Modules

The second key ingredient is D-finite functions. A function ϕ is said to be *differentiably finite* or just simply *D-finite* in $\mathbb{K}[[x_1, \dots, x_r]]$ if and only if the $\partial_1^{\alpha_1} \cdots \partial_r^{\alpha_r}\phi$ generate a finite dimensional vector space over $\mathbb{K}(x_1, \dots, x_r)$. In this case, ϕ is determined by a system of linear differential equations.

There are algorithms to make effective the closure of D-finite functions under $+$, $\times$, derivation, algebraic substitution, indefinite and definite integration, hadamard product and diagonals.

In order to treat symmetric functions, however, we first need to expand this definition to functions with an infinite number of variables. The function $\phi(x_1, x_2, \dots)$ is D-finite in $K[[x_1, x_2, \dots]]$ if for any r, $\phi(x_1, \dots, x_r, 0, \cdots)$ is D-finite in $K[[x_1, x_2, \dots, x_r]]$. This case does not enjoy all of the closure properties of the previous, nonetheless we have closure under $+$, $\times$, ∂_i, extension of coefficients, rational substitution, and exponentials of polynomials. We say that a symmetric function $\phi \in \hat{\Lambda}$ is D-finite if it is D-finite in $\mathbb{Q}[[p_1, p_2, \dots]]$. For example, the famous sums of symmetric functions, $\sum_n h_n = \prod_{i\geq 1}(1 + x_i) = \exp\left(\sum_n p_n/n\right)$ is D-finite under this definition.

In the symmetric case we also have closure under *Kronecker* (or tensor) product ($p_\lambda * p_\mu = \delta_{\lambda,\mu} z_\lambda p_\lambda$), and closure under plethysm with a polynomial. The property we are most interested here is that if ϕ and ψ are both D-finite symmetric functions in $K[t_1, t_2, \dots, t_n]$, and ϕ is a function of at most a finite number of power sum symmetric functions then $\langle \phi, \psi\rangle$ is a D-finite function of $k[t_1, t_2, \dots, t_n]$ [2].

2.1. The Weyl algebra.

We will present and prove our algorithm in the context of the *Weyl algebra* A_p of linear differential operators. This algebra is particularly useful for manipulating differential equations with polynomial coefficients. It is in this algebra that we can make many D-finite closure properties effective. We define $A_p := \mathbb{C}\langle p_1, \dots, p_n, \partial_1, \dots, \partial_n; \mathcal{R}_p\rangle$, with $\partial_i = d/dp_i$ and relations $\mathcal{R}_p : \partial_i p_j = p_j \partial_i + \delta_{i,j}$, and $p_i p_j - p_j p_i = \partial_i \partial_j - \partial_j \partial_i = 0$. We likewise consider additional variables, for example t and ∂_t, in $A_{p,t}$.

We shall denote left ideals and modules respectively by $I_\phi = \operatorname{ann}_{A_{p,t}} \phi = \{\, L \in A_{p,t} \mid L\phi = 0 \,\}$ and $A_{p,t}\phi = \{\, L\phi \mid L \in A_{p,t} \,\} \simeq A_{p,t}/I_\phi$. We filter A_p by total degree. First define $F_d = \{\, P \in A_p \mid \deg P \leq d \,\}$. Clearly, $F_d F_{d'} \subset F_{d+d'}$ $F_d \subset F_{d+1}$, and $\bigcup_{d\geq 0} F_d = A_p$. The module $A_p\phi$ is *holonomic* when $\dim_{\mathbb{C}} F_d\phi = O(d^n)$. The function ϕ is *holonomic* if $A_p\phi$ is holonomic. There is a key theorem of Kashiwara and Takayama which says that ϕ is holonomic if and only if ϕ is D-finite.

3. The Scalar Product of Symmetric Functions

3.1. Adjunction.

The magic operation that allows us to manipulate the scalar product is its adjunction, or adjoint operator. For P and Q in A_p, $\langle P\phi, \psi\rangle = \langle \phi, P^{\text{adj}}\psi\rangle, (P^{\text{adj}})^{\text{adj}} = P$, and

$(PQ)^{\mathrm{adj}} = Q^{\mathrm{adj}} P^{\mathrm{adj}}$. So for the *usual* scalar product of functions, $(f,g) \mapsto \int fg$, $\langle \partial_i \phi, \psi \rangle = -\langle \phi, \partial_i \psi \rangle$ implies that $p_i^\star = p_i$ and $\partial_i^\star = -\partial_i$, a bi-product of integration by parts. It is easy to calculate that in the case of the symmetric scalar product:

$$\langle p_i \phi, \psi \rangle = \langle \phi, i\partial_i \psi \rangle \quad \text{implies} \quad p_i^\diamond = i\partial_i \quad \text{and} \quad \partial_i^\diamond = i^{-1} p_i.$$

If M is a left module, then M^{adj} is a right module. These two modules have the same support and addition, however, we define multiplication this way for any$P \in A_p$, and any $m \in M^{\mathrm{adj}}$, $mP := (P^{\mathrm{adj}})m$. So, if $M = A_p\phi$, $M^{\mathrm{adj}} = \phi^{\mathrm{adj}} A_p$ with $\phi^{\mathrm{adj}} = \phi$, then $I_{\phi^{\mathrm{adj}}} = (I_\phi)^{\mathrm{adj}}$, and $(A_p\phi)^{\mathrm{adj}} \simeq A_p / I_{\phi^{\mathrm{adj}}}$.

3.2. **Calculating with the scalar product.** The algorithm also works in the more general cases, but for the sake of clarity and its good sister brevity we consider a unique auxiliary t variable, and we impose $\partial_t \phi = 0$.

The important link between annihilators of ϕ and ψ and of $\langle \phi, \psi \rangle$ is the following equation. Imagine a $W = (P^\diamond SU + TQ)$ with $S \in A_p$, $T \in A_{p,t}$, $U \in A_t$, $P \in I_\phi \cap A_p$ and $Q \in I_\psi$. Then,

$$\big\langle \phi, W\psi \big\rangle = \big\langle \phi, (P^\diamond SU + TQ)\psi \big\rangle = \langle S^\diamond P\phi, U\psi \rangle + \langle \phi, TQ\psi \rangle = \langle S^\diamond \cdot 0, U\psi \rangle + \langle \phi, T \cdot 0 \rangle = 0.$$

If by chance, $W \in \mathbb{C}\langle t, \partial_t \rangle$, since $\partial_t \langle \phi, \psi \rangle = \langle \phi, \partial_t \psi \rangle$, and $t\langle \phi, \psi \rangle = \langle t\phi, \psi \rangle = \langle \phi, t\psi \rangle$, then in fact, $W\langle \phi, \psi \rangle = \big\langle \phi, W\psi \big\rangle = 0$. So, if we calculate generators for $\big((\mathrm{ann}_{A_p}\, \phi)^\diamond A_t + (\mathrm{ann}_{A_{p,t}}\, \psi)\big) \cap \mathbb{C}\langle t, \partial_t \rangle$ any element in this intersection will annihilate the scalar product, that is, will represent a differential equation satisfied by the scalar product, which is PRECISELY WHAT WE WANT! Now, we must add that not all annihilators of the scalar product have to be of this form. However, that there are any is sufficient for our purposes.

The idea of the calculation is to iteratively construct elements in truncations of $(I_\phi)^\diamond$ and I_ψ, and at each step search for some linear combination which is in the ring A_t (for example, using our good friend, Gröbner bases.)

3.3. **Algorithm.** The inputs are ϕ D-finite in $\mathbb{C}[[p_1, \ldots, p_n]]$, and ψ D-finite in $\mathbb{C}[[t, p_1, \ldots, p_n]]$.

1. Calculate the Gröbner bases $(\mathcal{G}_\phi)^\diamond$ for $(\mathrm{ann}\, \phi)^\diamond$ and $\mathcal{G}_\psi$ for $(\mathrm{ann}\, \psi)$ with respect to the same order.
2. set $B = \{\}$;
3. iterate over monomials α from $[p_1, \ldots, p_n, \partial_1, \ldots, \partial_n, \partial_t]$ in an increasing order;
   (a) set α_ϕ to the remainder from *left* division of α by $(\mathcal{G}_\phi)^\diamond$;
   (b) set α_ψ to the remainder from *right* division of α by $\mathcal{G}_\psi$;
   (c) put both $\alpha - \alpha_\phi$ and $\alpha - \alpha_\psi$ in B and reduce to eliminate the p_i and the ∂_i by linear algebra;
   (d) If B contains an element P in t and ∂_t only, stop and output this element.

3.4. **Why does it work?** We build a helper intermediate space $S := (A_{p,t}\phi)^\diamond \otimes_{A_p[t]} (A_{p,t}\psi)$, which has a natural injective map onto $\langle A_{p,t}\phi, A_{p,t}\psi \rangle$, and in which we have the relations $(P\phi)^\diamond \otimes \psi = (\phi^\diamond P^\diamond) \otimes \psi = \phi^\diamond \otimes (P^\diamond \psi)$. In fact, $S \simeq \big(A_{p,t}^\diamond \otimes_{A_p[t]} A_{p,t}\big)/K$ for $K = I_\phi^\diamond \otimes_{A_p[t]} A_{p,t} + A_{p,t} \otimes_{A_p[t]} = I_\psi\, \mathrm{ann}_{A_{p,t}^\diamond \otimes_{A_p[t]} A_{p,t}} (\phi^\diamond \otimes \psi)$. The algorithm progressively constructs a basis for the vector space K, and then using linear algebra determines a basis for $K \cap \mathbb{C}\langle t, \partial_t \otimes 1 + 1 \otimes \partial_t \rangle$. This maps to $\big((\mathrm{ann}_{A_p}\, \phi)^\diamond A_t + (\mathrm{ann}_{A_{p,t}}\, \psi)\big) \cap \mathbb{C}\langle t, \partial_t \rangle$.

3.5. **Will it finish?** To prove termination, we first prove that S is infact a holonomic module whenever ϕ and ψ are holonomic. Essentially, this is done by building a module which injectively maps onto it, and the holonomy of this module is easily established, in part by passing through the usual scalar product. Since S is a holonomic module, the intersection $K \cap \mathbb{C}\langle t, \partial_t \otimes 1 + 1 \otimes \partial_t \rangle$ cannot reduce to $\{0\}$ because of a restriction on the dimension of S. The algorithm eventually produces every element of K, and thus the terminating condition will be satisfied.

4. **Applications to Combinatorics**

FINALLY, we return to the enumeration problem. Basically, we encode all graphs by their degree sequence. The sum of this encoding over all graphs, ϕ_G, gives a particularly nice D-finite symmetric series!

$$\phi_G = \sum_{g\in G} \prod_{\{i,j\}\in g} x_i x_j = \prod_{i<j}(1+x_i x_j) = \left(\exp\sum_{i\geq 1}\frac{p_i}{i}\right)[e_2].$$

To determine the number $r_{n,k}$ of graphs on n vertices with all valencies equal to k, we EXTRACT THE COEFFICIENT of $x_1^k x_2^k \cdots x_n^k$ in this giant series.

$$R_k(t) := \sum_{n\geq 0} r_{n,k}\frac{t^n}{n!} = \sum_{k\geq 0}\big\langle \phi_G, h_{\langle k^n\rangle}\big\rangle \frac{t^n}{n!} = \big\langle \phi_G, \sum_{n\geq 0} h_{\langle k^n\rangle}\frac{t^n}{n!}\big\rangle = \big\langle \phi_G, \exp(h_k t)\big\rangle.$$

This gives a scalar product that we can apply the algorithm to:

$$R_k(t) = \left\langle \exp\Bigg(\sum_{\substack{i\leq k\\ i \text{ pair}}} (-1)^{i/2}\frac{p_i^2}{2i} + \frac{p_i}{i} + \sum_{\substack{i\leq k\\ i \text{ impair}}}\frac{p_i^2}{2i}\Bigg), \exp\left(t\sum_{\lambda\vdash k}\frac{p_\lambda}{z_\lambda}\right)\right\rangle.$$

The current implementation of the algorithm can determine the differential equation satisfied by $R_k(t)$ for $t = 1, \ldots, 4$.

We can do a similar treatment to get $Y_k(t)$. Once this is done for a few values of k, the following conjecture arises.

Conjecture 1. *The number $y_{n,k}$ of k-uniform Young Tableaux of size kn has asymptotic growth like*

$$y_{n,k} \sim \frac{1}{\sqrt{2}}\left(\frac{e^{k-2}}{2\pi}\right)^{k/4} n!^{k/2-1}\left(\frac{k^{k/2}}{k!}\right)^n \frac{\exp(\sqrt{kn})}{n^{k/4}}, \qquad n\to\infty.$$

The scalar product is closely related to the Kronecker product and using a variation of the algorithm we have the following.

Proposition 1. *The following identity holds, where s_λ is the Schur function indexed by λ,*

$$\left(\sum_\lambda s_\lambda\right) * \left(\sum_\lambda s_\lambda\right) = \exp\left(\sum_{n\geq 1}\frac{p_{2n-1}}{(2n-1)(1-p_{2n-1})}\right)\left(\prod_{n\geq 1}(1-p_n^2)\right)^{-1/2}.$$

Bibliography

[1] Chyzak (Frédéric), Mishna (Marni), and Salvy (Bruno). – Effective scalar products of D-finite symmetric functions. *Journal of Combinatorial Theory, Series A*, 2005. – In press.

[2] Gessel (Ira M.). – Symmetric functions and P-recursiveness. *Journal of Combinatorial Theory, Series A*, vol. 53, n° 2, 1990, pp. 257–285.

[3] Stanley (Richard P.). – *Enumerative combinatorics. Vol. 2.* – Cambridge University Press, Cambridge, 1999, *Cambridge Studies in Advanced Mathematics*, vol. 62, xii+581p.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 11–12.

Available online at the URL
http://algo.inria.fr/seminars/.

Heaps of Segments and Lorentzian Quantum Gravity

Will James

Department of Mathematics and Statistics, University of Melbourne (Australia)

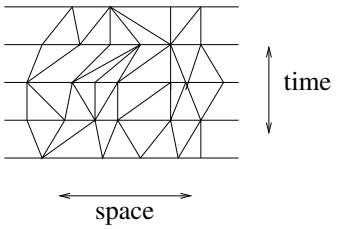
December 15, 2003

Summary by Sylvie Corteel

Abstract

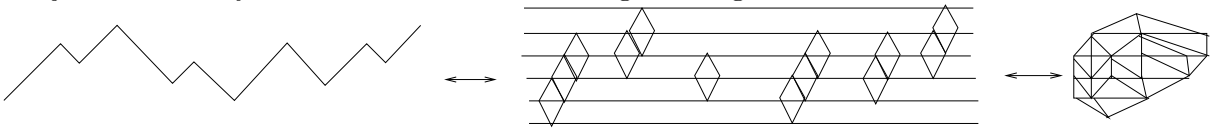
This work is a combinatorial study of quantum gravity models related to Lorentzian quantum gravity. These models are discrete combinatorial objects called dynamical triangulations. They are related to classical combinatorial objects: Dyck paths, heaps, ... This work is in collaboration with Xavier Viennot and is part of the speaker's thesis [4].

Quantum gravity is a quantum description of the space-time geometry. We refer to Loll [5] for precise definitions. In the Lorentzian quantum gravity the universe can be of dimension (1+1) [1]. One dimension is for space, the other one for time. This defines the dynamical triangulations.

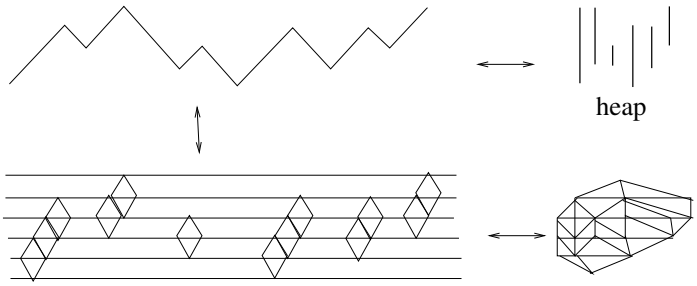


Moreover space is supposed to be circular and to have a unique origin. This gives some special cylindric dynamical triangulations (CDT). These CDT can be cut and made planar by choosing a first vertex following its rightmost link and cutting. These gives dynamical triangulations with the condition that all the rightmost triangles points up.

The problem is therefore to enumerate these triangulations according to different parameters, for example according to the number of triangles. The model was solved by Di Francesco, Guitter and Kirstjansen [3] and they included a parameter that measured the absolute value of the local curvature. They used transfer matrix methods to solve the model, which they presented in terms of a bijection with Dyck Paths that end at their highest height.



This bijection is the motivation for this combinatorial study of quantum gravity models. The authors show an equivalent bijection between CDTs and heaps of segments. The number of triangles and the curvature are easy to read on the heap. The number of triangles are the sum of the width of the segments and the curvature is twice the number of segments minus the number of contacts. We refer to [4] for detailed definitions. Then using the heap machinery [7, 6] multi-variate generating functions can be deduced.



This gives a series of solutions of existing models and show how different restrictions to the topology of the universes can be modelled by similar classes of heaps. In particular connected heaps [2] appear naturally. This work and extensions can be found in the speaker's thesis [4].

Bibliography

[1] Ambjorn (J.) and Loll (R.). – Non-perturbative Lorentzian quantum gravity. *Nuclear Physics B*, vol. 536, (1998), pp. 407–434.

[2] Bousquet-Mélou (M.) and Rechnitzer (A.). – Lattice animals and heaps of dimers. *Discrete Mathematics*, vol. 258, n° 1-3, (2002), pp. 235–274.

[3] Francesco (P. Di), Guitter (E.), and Kristjansen (C.). – Integrable 2D Lorentzian gravity and random walks. *Nuclear Physics B*, vol. 567, n° 3, 2000, pp. 515–553.

[4] James (Will). – Phd thesis. *University of Melbourne*, (2005).

[5] Loll (R.). – A discrete history of the Lorentzian path integral. *Lecture Notes in Physics*, vol. 631, (2003), pp. 137–171.

[6] Viennot (Gérard Xavier). – Heaps of pieces. I. Basic definitions and combinatorial lemmas. In *Combinatoire énumérative (Montreal, Que., 1985/Quebec, Que., 1985)*, pp. 321–350. – Springer, Berlin, 1986.

[7] Viennot (Gérard Xavier). – Heaps of pieces. I. Basic definitions and combinatorial lemmas. In *Graph theory and its applications: East and West (Jinan, 1986)*, pp. 542–570. – New York Academy of Sciences, New York, 1989.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 13–20.

Available online at the URL
http://algo.inria.fr/seminars/.

Matrix Models an Knot Theory

Paul Zinn-Justin
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud 11 (France)

January 26, 2004

Summary by Dominique Gouyou-Beauchamps

Abstract

We shall explain how knot, link and tangle enumeration problems can be expressed as matrix integrals which will allow us to use quantum field-theoretic methods. We shall discuss the asymptotic behaviors for a great number of intersections. We shall detail algorithms used to test our conjectures.

1. Classification and Enumeration of Knots, Links, Tangles

A knot is defined as a closed, non-self-intersecting curve that is embedded in three dimensions and cannot be untangled to produce a simple loop (i.e., the unknot). A knot can be represented by its plane projection (i.e., its diagram). A knot can be generalized to a link, which is simply a knotted collection of one or more closed strands. A tangle is defined as a region in a knot or link projection plane surrounded by a circle such that the knot or link crosses the circle exactly four times. An alternating knot (resp. link) is a knot (resp. link) which possesses a knot diagram (resp. link diagram) in which crossings alternate between under- and overpasses (see Figure 1).

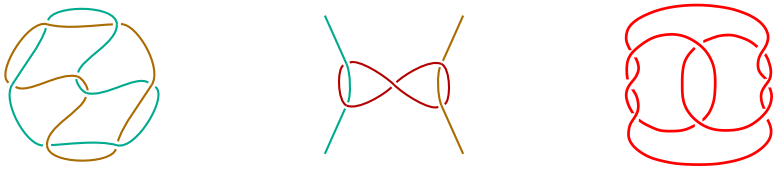


FIGURE 1. An alternating link, a tangle and an non-alternating knot

FIGURE 2. A 6_1 knot of the Tait's classification

P. G. Tait [16, 17, 18, 19, 20, 21] undertook a study of knots in response to Kelvin's conjecture that the atoms were composed of knotted vortex tubes of ether (Thompson [24]). He categorized knots in terms of the number of crossings in a plane projection (see Figure 2). He also made some conjectures which remained unproven until the discovery of Jones polynomials:

1. Reduced alternating diagrams have minimal link crossing number,
2. Any two reduced alternating diagrams of a given knot have equal writhe,
3. The flyping conjecture, which states that the number of crossings is the same for any reduced diagram of an alternating knot (see [25] for definition of the flyping equivalence).

Conjectures (1) and (2) were proved by Kauffman [4], Murasugi [9], and Thistlethwaite [22, 23] using properties of the Jones polynomial or Kauffman polynomial **F** (see Hoste et al. [1]). Conjecture (3) was proved true by Menasco and Thistlethwaite [7, 8] using properties of the Jones polynomial. Schubert [12] showed that every knot can be uniquely decomposed (up to the order in which the decomposition is performed) as a knot sum of a class of knots known as prime knots, which cannot themselves be further decomposed. Knots that are the sums of prime knots are known as composite knots.

There is no known formula for giving the number of distinct prime knots as a function of the number of crossings. The numbers of distinct prime knots having $n = 1, 2, \ldots$ crossings are 0, 0, 1, 2, 3, 7, 21, 49, 165, 552, 2176, 9988, ... (Sloane's M0851) [13].

In the 1932, Reidemeister [10] first rigorously proved that knots exist which are distinct from the unknot. He did this by showing that all knot deformations can be reduced to a sequence of three types of "moves," called Reidemeister moves (see Figure 3).

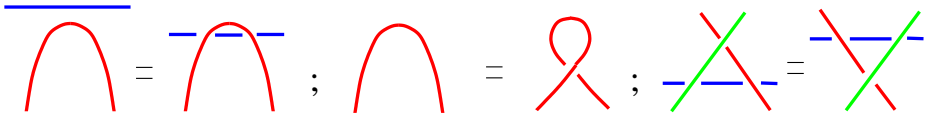


FIGURE 3. The three Reidemeister moves (poke, twist and slide)

2. Feynman Diagrams. $O(n)$ Matrix Model and Renormalization

We want to enumerate prime alternating tangles with a given number of components and crossing. Let $a_{k;p}$ be the number of prime alternating tangles with $k+2$ components and p crossings. Let $\Gamma(n, g)$ be the corresponding generating series:

$$\Gamma(n, g) = \sum_{k=0, p=1}^{\infty} a_{k;p}\, n^k g^p$$

In [29], we have shown that the integral

$$Z = \int dM e^{N[-\frac{1}{2} trM^2 + \frac{g}{4} trM^4]} \quad \text{where} \quad dM = \prod_i dM_{ii} \prod_{i<j} d\Re eM_{ij}\, d\Im mM_{ij} \tag{1}$$

over $N \times N$ hermitean matrices is well suited for the counting of alternating links and tangles: for an appropriate choice of $\alpha(g)$, $2\frac{\partial}{\partial g} \lim_{N\to\infty} \frac{1}{N^2} \log Z_n(g, \alpha(g))$ is the generating function of the number of alternating tangle diagrams with n 4-valent crossings. In the context of knot theory, it seems natural to consider this integral over complex (non hermitean) matrices, in order to distinguish between under-crossing and over-crossing. However, this integral is closely related to the simpler integral (1) in the large N limit.

It has been known to physicists since the pioneering work of 't Hooft [15] that the large N limit of the previous integral (1) may be organized in a topological way. While the leading term corresponds to planar diagrams, the subdominant terms of order N^{-2h} of $\frac{1}{N^2} \log Z(g)$ are describe by graphs drawn on a Riemann surface of genus h.

The series expansion of Z in powers of g may be represented diagrammatically by Feynman diagrams, made of undirected edges or "propagators" with double lines expressing the conservation of indices, $\langle M_{ij}M_{kl}\rangle_0 = \frac{1}{N}\delta_{il}\delta_{jk}$ and the 4-valent vertices $gN\delta_{qi}\delta_{jk}\delta_{lm}\delta_{np}$ (see Figure 4). More precisely

$$\lim_{n\to\infty}\frac{1}{N^2}\log Z = \sum \text{ weight } g^n$$

where the sum is over all planar diagrams with n vertices and with a weight equal one over the order of the automorphism group of the diagram.

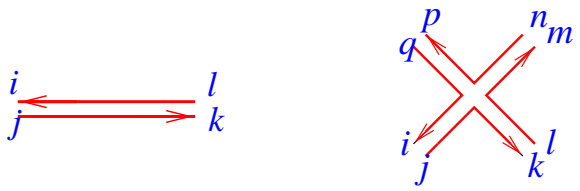


FIGURE 4. Propagator and 4-valent vertex

In order to connect integral (1) with knot theory, we take any planar diagram (i.e., 4-regular planar map) and to the following: starting from an arbitrary crossing, we decide it is a crossing of two strings (again there is an arbitrary choice of which is under/over-crossing). Once the first choice is made, we simply follow the string and form alternating sequences of under- and over-crossings. The remarkable fact is that this can be done consistentlty (see Figure 5). If we identify two alternating diagrams obtained from one another by inverting undercrossings and overcrossings, then there is a one-to-one correspondence between planar diagrams and alternating link diagrams.

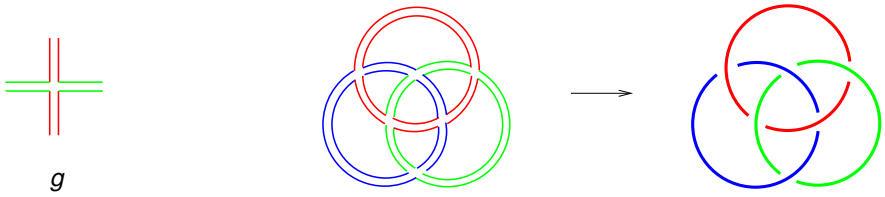


FIGURE 5. A planar diagram and one of the two corresponding alternating link diagrams

In order to distinguish strands of links, we introduce a more general model, which we shall call the intersecting loops $O(n)$ model. If n is a positive integer, then consider the following multi-matrix integral:

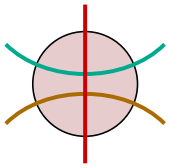
$$Z^{(N)}(n,g) = \int\prod_{a=1}^{n} d\,M_a\, e^{N\,tr\left(-\frac{1}{2}M_a^2+\sum_{b=1}^{n}\frac{g}{4}(M_aM_b)^2\right)} \tag{2}$$

and the corresponding free energy

$$F(n,g) = \lim_{N\to\infty}\frac{\log Z^{(N)}(n,g)}{N^2} = \sum_{k,p=1}^{\infty} f_{k;p}n^k g^p.$$

The correlation functions count tangle diagrams:

$$\lim_{N\to\infty}\left\langle \frac{1}{N} tr(M_1M_2M_3M_2M_1M_3)\right\rangle_c =$$



.

For tangles, a convenient way to keep track of the number of connected components and of the connections of the external legs is to use colors. The colors allow us to distinguish the various external legs and add an extra power series variable in the theory (the number of colors n) to count separately objects with different numbers of connected components.

This model has two problems:

1. the diagrams generated by applying Feynman rules are not necessarily reduced or prime,
2. several reduced diagrams may correspond to the same knot due to the flyping equivalence.

We are therefore led to renormalize the quadratic and quartic interaction of (2). A key observation is that, while there is only one such quadratic $O(n)$-invariant term, there are two quartic $O(n)$-invariant terms, which leads to a generalized model with 3 coupling constants (i.e., t, g_1 and g_2) in the action (bare coupling constants):

$$Z^{(N)}(n,t,g_1,g_2) = \int \prod_{a=1}^{n} d\,M_a e^{Ntr\left[-\frac{t}{2}M_a^2+\sum_{b=1}^{n}\left(\frac{g_1}{4}M_aM_bM_aM_b+\frac{g_2}{2}M_aM_aM_bM_b\right)\right]} \tag{3}$$

where t, g_1 and g_2 are functions of the renormalized coupling constant g, chosen such that the correlation functions are the appropriate generating series in g of the numbers of alternating links (see [28] and Figure 6).

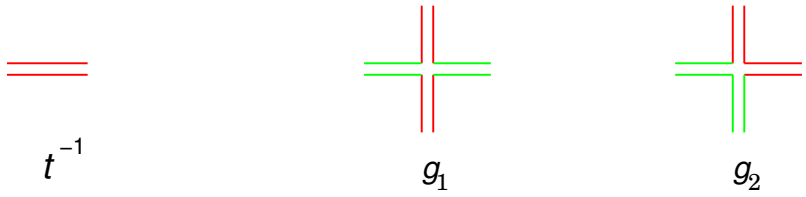


FIGURE 6. The three coupling constants

There are currently two values of n for which the corresponding matrix model has been exactly solved: $n = 1$ and $n = 2$.

The case $n = 1$ is particulary important since it corresponds to counting all alternating tangles regardless of the number of connected components. We have the usual matrix model

$$Z^{(N)}(t,g) = \int dM\, e^{N\,tr\left(-\frac{t}{2}M^2+\frac{g}{4}M^4\right)} \quad \text{with } g = g_1 + 2g_2\ .$$

"Renormalization" equations recombine into a fifth degree equation:

$$32 - 64\,A + 32\,A^2 - 4\frac{1+2g-g^2}{1-g}A^3 + 6\,g\,A^4 - g\,A^5 = 0.$$

Correlation functions are given in terms of its solution. In particular, if $\left\langle \frac{1}{N} trM^{2\ell}\right\rangle_c = \sum_{p=0}^{\infty} a_p g^p$ is the generating function of prime alternating tangles with 2ℓ legs where p counts the crossings, then

$$a_p \overset{p\to\infty}{\sim} \text{cst}\, g_c^{-p}\, p^{-5/2} \quad \text{with } g_c = \frac{\sqrt{21001}-101}{270}\ (g_c^{-1} \approx 6.147930)\ .$$

These works generalize known results: the case $\ell = 2$ (Sundberg and Thistlethwaite [14]), and Zinn-Justin and Zuber [29]. The number f_p of prime alternating links grows like $f_p \sim \mathrm{cst}\, g_c^{-p} p^{-7/2}$ (Kunz-Jacques and Schaeffer [6]).

In case $n = 2$, integral (2) is equivalent to integral (4) that is recently studied in detail and computed in the framework of the random lattice model (Zinn-Justin [26] and Kostov [5]). In [30], we thus carry out the explicit counting of alternating 2-color tangles: their generating function is the solution of coupled equations involving elliptic functions.

$$(4) \qquad Z^{(N)}(t,g_1,g_2) \;=\; \int dM_1\, dM_2\, e^{N\, tr\left[-\frac{t}{2}(M_1^2+M_2^2)+\frac{g_1+2g_2}{4}(M_1^4+M_2^4)+\frac{g_1}{2}(M_1M_2)^2+g_2M_1^2M_2^2\right]}$$

When we introduce a complex matrix $X = \frac{1}{\sqrt{2}}\,(M_1 + iM_2)$, we obtain:

$$Z^{(N)}(t,b,c) = \int dX\, dX^\dagger\, e^{N\, tr\left(-tXX^\dagger + bX^2X^{\dagger 2} + \frac{1}{2}c(XX^\dagger)^2\right)} \quad \text{with } b = g_1+g_2 \text{ and } c = 2g_2\ .$$

The number γ_p of prime alternating 2-color tangles with p crossings grows like

$$\gamma_p \overset{p\to\infty}{\sim} \mathrm{cst}\, g_c^{-p} p^{-2} (\log p)^{-1} \qquad \text{with } g_c^{-1} \approx 6.28329764\ .$$

In [27], we establish that the number f_p of reduced alternating link diagrams with two colors and p crossings has the following asymptotics:

$$f_p \overset{p\to\infty}{\sim} \mathrm{cst}\, g_c^{-p}\; p^{-3} \log p \qquad \text{with } g_c = \frac{\pi(\pi-4)^2}{16}\ (g_c^{-1} \approx 6.91167)\ .$$

The number $g_c^{-1} = 6.91167\ldots$ is slightly larger than the value 6.75 obtained for only one color.

3. **Algorithm: Transfer Matrix**

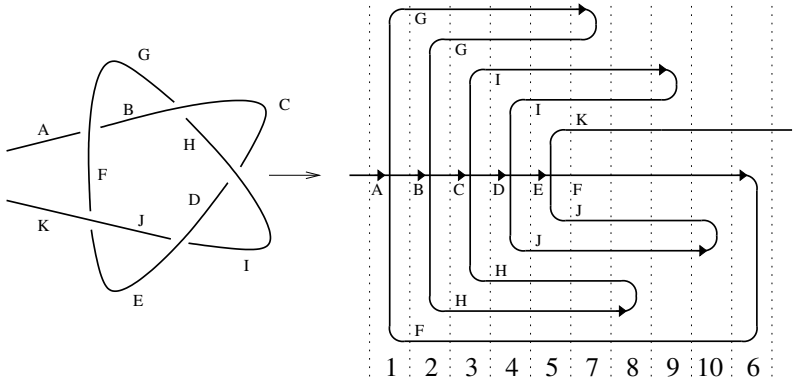


FIGURE 7. The steps are ordered by their number below the diagram. At each step, the active line is distinguished by an arrow.

In [2, 3], we propose a new method to enumerate alternating knots using a transfer matrix approach. The basic ingredient is the ability to cut the object one is studying into slices, which represent the state of the system a fixed (discrete) time. The naive idea would be to draw the knot diagrams on the plane in such a way that time would correspond to one particular coordinate of the plane, that is to read the knot diagrams "from left to right." Here, this idea does not work directely, and one is led to a slightly more sophisticated notion of slices, which we shall explain using the example of Figure 7.

A basis state will be described by a series of left and right arches and the position of the active line. As an illustration, we show all the intermediate states of the example of Figure 7 on Figure 8.

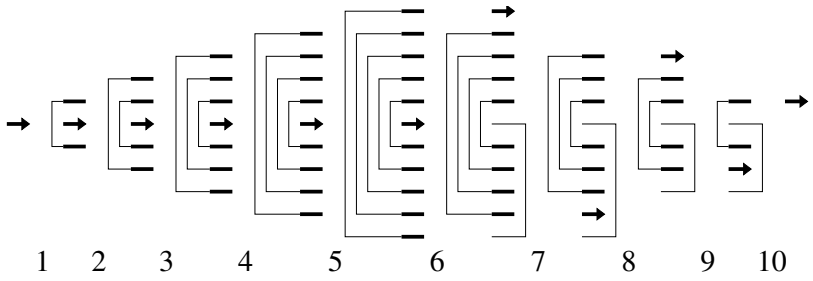


FIGURE 8. A sequence of intermediate states. The active line is denoted by an arrow.

The table below gives numbers $a_{k,p}$ of prime alternating tangles with 2 external legs, k circles (i.e., $k+1$ connected components) and p crossings up to $p = 15$ (even though they can be easily obtained for p up to 18 or 19, as in [3], on a work station, and probably further using larger computers). Tangles of types 1 (i.e., Γ_1) and 2 (i.e., Γ_2) are distiguished by the two ways of connecting their external legs. The reader is reminded that the total number of tangles is given by $\Gamma_1 + 2\Gamma_2$.

	Γ_1							Γ_2						
p \ k	0	1	2	3	4	5	6	0	1	2	3	4	5	6
1	1							0						
2	0							1						
3	2							1						
4	2							3	1					
5	6	3						9	1					
6	30	2						21	11	1				
7	62	40	2					101	32	1				
8	382	106	2					346	153	24	1			
9	1338	548	83	2				1576	747	68	1			
10	6216	2968	194	2				7040	3162	562	43	1		
11	29656	11966	2160	124	2			31556	17188	2671	121	1		
12	131316	71422	9554	316	2			153916	80490	15295	1484	69	1	
13	669138	328376	58985	5189	184	2		724758	425381	87865	6991	194	1	
14	3156172	1796974	347038	22454	478	2		3610768	2176099	471620	52231	3280	103	1
15	16032652	9298054	1864884	193658	10428	260	2	17853814	11376072	2768255	308697	15431	290	1

4. **Algorithm: Random Sampling**

In [11], in order to generate a random map according to the uniform distribution on rooted 4-regular planar maps with p vertices one generate a blossom tree according to the uniform distribution on blossom tree and apply closure that is a $(p+2)$-to-2 correspondence between blossom trees and rooted 4-regulat maps (see Figure 9). This algorithm allows to generate in linear time (up to $p = 10^7$ vertices) rooted 4-regular planar maps with p vertices and two legs, with uniform probability. One can compute various quantities related to the map thus generated and then average over a sample, as always in Montecarlo simulations.

The main idea of the physical interpretation of the number $a_p(1)$ of rooted 4-regular maps is to consider planar maps as discretized random surfaces (with the topology of the sphere). As the

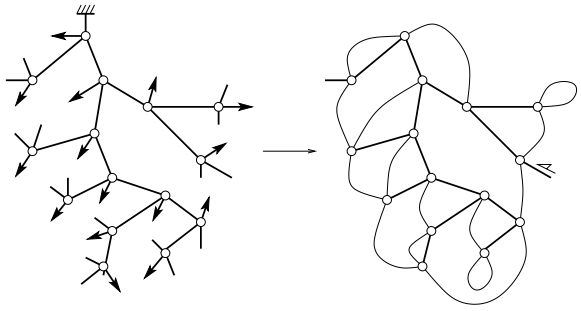


FIGURE 9. Schaeffer's bijection between blossom trees and planar maps.

number of vertices of the map grows large, the details of the discretization can be assimilated to the fluctuations of the metric on the sphere. Now, physics tell us that the metric is the dynamical field of general relativity, i.e., gravity, and that this type of fluctuations in the metric are characteristic of a quantum theory. In our case it means that, as p becomes large, the discrete nature of the maps can be ignored and there exists a scaling limit, the properties of which are described by two-dimensional euclidian gravity. In particular any parameter of random planar maps that makes sense in the scaling should converge to its continuum analog. A fundamental parameter of this kind turns out to be precisely the number of (unrooted) planar maps: it is expected to scale to the partition function $Z_g(A)$ of two-dimensional quantum gravity with spherical topology at fixed area A, through a relation of the form $\frac{1}{p}a_p(1) \overset{p\to\infty}{\sim} Z_g(A)$, with A proportional to p.

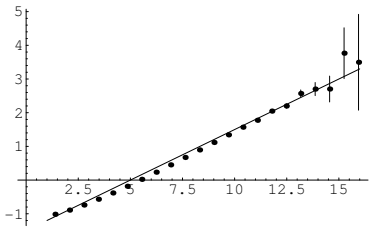


FIGURE 10

Here the factor $1/p$ is due to the fact that the partition function does not takes the rooting into account. We conjecture that for $|n| < 2$, the matrix model is in the universality class of a 2D field theory with spontaneously broken $O(n)$ symmetry, coupled to gravity. The large size limit is described by a a conformal field theory (CFT) coupled to gravity with $c = n - 1$:

$$a_p(n) \sim \mathrm{cst}(n)\, g_c(n)^{-p}\, p^{\gamma(n)-2}, \qquad f_p(n) \sim \mathrm{cst}(n)\, g_c(n)^{-p}\, p^{\gamma(n)-3},$$
$$\gamma = \frac{c - 1 - \sqrt{(1-c)(25-c)}}{12}.$$

In particular, knots correspond to the limit $n \to 0$:

$$f_p(0) \sim \mathrm{cst}\, g_c^{-p} p^{-\frac{19+\sqrt{13}}{6}}$$

With Schaeffer's algorithm, we have tested quantity: $\gamma' \equiv \frac{d\gamma}{dn}_{|n=1} = 3/10$ according to the conjecture. We obtain a very good agreement (see Figure 10).

Bibliography

[1] Hoste (J.), Thistlethwaite (M.), and Weeks (J.). – The first 1,701,936 knots. *Mathematical Intelligencer*, vol. 20, 1998, pp. 33–48.

[2] Jacobsen (J.-L.) and Zinn-Justin (P.). – A Tranfer Matrix approach to the Enumeration of Colored Links. *Journal of Knot Theory and its Ramifications*, vol. 10, 2001, pp. 1233–1267. – `http://arXiv.gov/math-ph/0104009`.

[3] Jacobsen (J.-L.) and Zinn-Justin (P.). – A Tranfer Matrix approach to the Enumeration of Knots. *Journal of Knot Theory and its Ramifications*, vol. 11, 2002, pp. 739–758. – `http://arXiv.gov/math-ph/0102015`.

[4] Kauffman (L.H.). – State models and the jones polynomial. *Topology*, vol. 26, 1987, pp. 395–407.

[5] Kostov (I.K.). – Exact solution of the six-vertex model on a random lattice. preprint `http://arXiv.gov/hep-th/9911023`.

[6] Kunz-Jacques (S.) and Schaeffer (G.). – The aymptotic number of prime alternating link. In Barcelo (H.) (editor), *Proceedings of the 13th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC'01)*. – Tempe, Arizona (USA), May $20 - 26$, 2001.
Available at the URL `http://www.lix.polytechnique.fr/Labo/Gilles.Schaeffer/Biblio/`.

[7] Menasco (W.) and Thistlethwaite (M.). – The wait flyping conjecture. *Bulletin of the American Mathematical Society*, vol. 25, 1991, pp. 403–412.

[8] Menasco (W.) and Thistlethwaite (M.). – The classification of alternating links. *Annals of Mathematics*, vol. 138, 1993, pp. 113–171.

[9] Murasugi (K.). – The jones polynomial and classical conjectures in knot theory. *Topology*, vol. 26, 1987, pp. 187–194.

[10] Reidemeister (K.). – *Knotentheorie*, pp. 273–347. – Verlag von Julius Springer, Berlin, 1932.

[11] Schaeffer (G.) and Zinn-Justin (P.). – On the Asymptotic Number of Plane Curves and Alternating Knots. – `http://arXiv.gov/math-ph/0304034`.

[12] Schubert (H.). – *Die eindeutige Zerlegbarkeit eines Knotens in Primknoten*, pp. 57–104. – H. Sitzungsber. Heidelberger Akad. Wiss., Math.-Natur. Kl., 1949, vol. 3.

[13] Sloane (N.J.A.) and Plouffe (S.). – *The Encyclopedia of Integer Sequences*. – Academic Press, Inc., 1995.

[14] Sundberg (C.) and Thistlethwaite (M.). – The Rate of Growth of the Number of Prime Alternating Links and Tangles. *Pacific Journal of Mathematics*, vol. 182, 1998, pp. 329–358.

[15] 't Hooft (G.). – A planar diagram theory for strong interactions. *Nuclear Physics*, vol. B 72, 1974, pp. 461–473.

[16] Tait (P.G.). – On knots. *Proceedings of the Royal Society of Edinburgh*, vol. 9, n° 97, $1876 - 7$, pp. 306–317.

[17] Tait (P.G.). – On knots I. *Transactions of the Royal Society of Edinburgh*, vol. 28, $1876 - 7$, pp. 145–190.

[18] Tait (P.G.). – On links. *Proceedings of the Royal Society of Edinburgh*, vol. 9, n° 98, $1876 - 7$, pp. 321–332.

[19] Tait (P.G.). – On knots II. *Transactions of the Royal Society of Edinburgh*, vol. 32, $1883 - 4$, pp. 327–342.

[20] Tait (P.G.). – On knots III. *Transactions of the Royal Society of Edinburgh*, vol. 32, $1884 - 5$, pp. 493–506.

[21] Tait (P.G.). – *On Knots I, II, and III. Scientific Papers*, pp. 273–347. – London: Cambridge University Press, 1898, vol. 1.

[22] Thistlethwaite (M.B.). – A spanning tree expansion of the jones polynomial. *Topology*, vol. 26, 1987, pp. 297–309.

[23] Thistlethwaite (M.B.). – Kauffman's polynomial and alternating links. *Topology*, vol. 27, 1988, pp. 311–318.

[24] Thompson (W.T.). – On vortex atoms. *Philo. Mag.*, vol. 34, 1867, pp. 15–24.

[25] Weisstein (E.W.). – Knot Theory. From MathWorld—A Wolfram Web Resource, `http://mathworld.wolfram.com/KnotTheory.html`.

[26] Zinn-Justin (P.). – The six-vertex model on random lattices. *Europhysics Letters*, vol. 50, 2000, pp. 15–21. – `http://arXiv.gov/cond-mat/9909250`.

[27] Zinn-Justin (P.). – Some Matrix Integrals related to Knots and Links. In *Proceedings of the 1999 semester of the MSRI "Random Matrices and their Apllications"*. – MSRI Publications Vol. 40, 2001. `http://arXiv.gov/math-ph/9910010`.

[28] Zinn-Justin (P.). – The General O(n) Quartic Matrix Model and its application to Counting Tangles and Links. *Communications in Mathematical Physics*, vol. 238, 2003, pp. 287–304. – `http://arXiv.gov/math-ph/0106005`.

[29] Zinn-Justin (P.) and Zuber (J.-B.). – Matrix Integrals and the Counting of Tangles and Links. In *Proceedings of the 11th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC'99)*. – Barcelona, 1999. Discrete Mathematics 246 (2002), 343-360, `http://arXiv.gov/math-ph/9904019`.

[30] Zinn-Justin (P.) and Zuber (J.-B.). – On the Counting of Colored Tangles. *Journal of Knot Theory and its Ramifications*, vol. 9, 2000, pp. 1127–1141. – `http://arXiv.gov/math-ph/0002020`.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 21–24.

Available online at the URL
http://algo.inria.fr/seminars/.

Particle Seas and Basic Hypergeometric Series

Sylvie Corteel
Prism, Université de Versailles–Saint-Quentin (France)

January 26, 2004

Summary by Julien Fayolle

Abstract

The author introduces overpartitions and particle seas as a generalization of partitions. Both new tools are used in bijective proofs of basic hypergeometric identities like the q-binomial theorem, Jacobi[1]'s triple product, q-Gauß[2] equality or even Ramanujan[3]'s ${}_1\Psi_1$ summation.

1. Partitions

In 1969, G. E. Andrews was already looking for bijective proofs for some basic hypergeometric identities. The principle of bijective proofs is simple: if each side of an equation can be construed as a generating function counting some parameters for sets A and B of combinatorial objects, and we can go bijectively between objects of the two sets transforming the parameters on objects from A into the parameters of objects from B, then we have an identity.

Let us first recall the notation

$$(a;q)_\infty := \prod_{n\geq 0}(1-aq^n) \quad \text{and} \quad (a;q)_k := \frac{(a;q)_\infty}{(aq^n;q)_\infty} = \prod_{n=0}^{k-1}(1-aq^n).$$

If no ambiguity arises we note $(a;q)_k = (a)_k$. For a partition $\lambda = (\lambda_1, \cdots, \lambda_k)$, we note $l(\lambda) = k$ its number of parts and $|\lambda| = \lambda_1 + \ldots + \lambda_k$ its weight. If nothing is specified, the partitions have positive parts (i.e., $\lambda_i > 0$). The generating functions for various types of partitions can be easily found in the literature (for example in [1])

The idea of introducing the particle seas stems from a bijective proof of Jacobi's triple product found by Itzykson (this proof was never published before [2]). We transform a usual partition into a *particle sea* (a precise and complete definition will be given later) by filling the zero-line with green balls and putting a blue ball in the column of ascissa x for any part x in the partition.

2. Overpartitions and the q-Binomial Theorem

In [3], Joichi and Stanton prove the q-binomial theorem through bijective means. We explain this bijection in terms of *overpartitions*, although the authors did not use this language.

Definition 1. An overpartition is a partition in which the *last* occurrence of a part *can* be overlined.

Remark. One can also define an overpartition with *first* instead of *last*.

[1]Karl Jacobi (1804–1851), German mathematician.
[2]Karl Friedrich Gauß (1777–1855), German mathematician, physicist, and astronomer.
[3]Srinivasa Aiyangar Ramanujan (1887–1920), Indian mathematician.

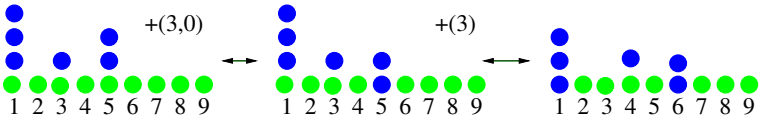


FIGURE 1. An example for Joichi and Stanton's bijection

For example, $(9,9,\overline{8},5,\overline{5},3,3,1,1,\overline{1})$ is an overpartition.

We first devise the generating function for the set O of overpartitions. We note that an overpartition can be decomposed bijectively in two parts: the overlined parts which make a partition into distinct parts (let D denote the set of partitions into distinct parts) and the non-overlined parts that are a general partition (denoted P). We introduce $o(\lambda)$ the number of overlined parts in the overpartition λ, and we have

$$\sum_{\lambda\in O} q^{|\lambda|}z^{l(\lambda)}x^{o(\lambda)} = \sum_{(\mu,\nu)\in D\times P} q^{|\mu|+|\nu|}z^{l(\mu)+l(\nu)}x^{l(\mu)} = \sum_{\mu\in D} q^{|\mu|}(zx)^{l(\mu)} \sum_{\nu\in P} q^{|\nu|}z^{l(\nu)} = (-qzx)_\infty \frac{1}{(zq)_\infty}.$$

For the reciprocal bijection one has to notice that the overlined parts are bound to be the last (or first, depending on the definition) ones with their respective weight.

We have to introduce additional notation: the set of overpartitions into k parts is $\mathcal{O}_k$, the set of partitions into distinct nonnegative parts smaller than k is $D_{k,\geq}$ and the set of partitions into parts less or equal to k is P_k.

Joichi and Stanton see an overpartition λ from $\mathcal{O}_k$ as the product of two partitions: one α from $P_k - P_{k-1}$ (i.e., the set of partitions into parts less or equal to k with at least one part of size k, this is in bijection with the set of partitions made of exactly k parts [it can be seen by rotating the partition diagram by $\pi/2$]), and one β from $D_{k,\geq}$. Here is the algorithm describing their bijection between $(P_k - P_{k-1}) \times D_{k,\geq}$ and $\mathcal{O}_k$:

- Transform α into a particle sea;
- For each part i in β, shift the first i blue balls in the sea by one to the right;
- Crash the $(i+1)$st ball on the zero-line (of the same column), destroying the green ball underneath it.

It follows that the partition λ has the characteristics $|\lambda| = |\alpha| + |\beta|$, $l(\lambda) = l(\alpha)$ and $o(\lambda) = l(\beta)$.

For the example in Figure 1 with $\alpha = (5,5,3,1,1,1)$, for $k = 6$ parts, and $\beta = (3,0)$. (We actually take a partition from the set of partitions in k parts but since this set is in bijection with $P_k - P_{k-1}$ we are OK.) We then have a sea particle that can be construed as an overpartition: For any blue ball on the zero-line, its abscissa is an overlined part in the overpartition, the other blue balls add a part in the overpartition equal to their abscissa. We obtain an overpartition into $k = 6$ positive parts $\lambda = (\overline{6}, 6, 4, \overline{1}, 1, 1)$.

This bijection means that

$$\sum_{\lambda\in\mathcal{O}_k} q^{|\lambda|}x^{o(\lambda)}z^{l(\lambda)} = \sum_{(\alpha,\beta)\in(P_k-P_{k-1})\times D_{k,\geq}} q^{|\alpha|+|\beta|}x^{l(\beta)}z^k = \sum_{\alpha\in P_k-P_{k-1}} q^{|\alpha|}z^k \sum_{\beta\in D_{k,\geq}} q^{|\beta|}x^{l(\beta)}$$

$$= z^k \left(\sum_{\alpha\in P_k} q^{|\alpha|} - \sum_{\alpha\in P_{k-1}} q^{|\alpha|} \right) (-x)_k = \frac{z^k q^k}{(q)_k}(-x)_k.$$

And from here we have

$$\sum_{\lambda\in O} q^{|\lambda|}x^{o(\lambda)}z^{l(\lambda)} = \sum_{k\geq 0}\sum_{\lambda\in\mathcal{O}_k} q^{|\lambda|}x^{o(\lambda)}z^{l(\lambda)} = \sum_{k\geq 0}\frac{z^k q^k(-x)_k}{(q)_k} = \frac{(-qzx)_\infty}{(zq)_\infty},$$

which is the q-binomial theorem.

Zeilberger also described a bijection to prove the q-binomial theorem that can be translated in the language of overpartitions.

3. Particle Seas

A particle sea can be interpreted as a couple of overpartitions with one overpartition into non-negative parts.

Definition 2. A particle sea is the upper half discrete plane partly filled with squares and balls such that:

- The zero-line is partially filled with squares and balls
- The rest of the positive quarter is partially filled with squares
- The rest of the non-positive quarter is partially filled with balls
- If the point (x, y) with $y > 0$ is filled, then the point $(x, y-1)$ is also filled
- If there is a ball (resp. square) of positive (resp. non-positive) abscissa then there is a square right (resp. ball left) to it.

Here the balls are graphically represented as green balls and the squares as blue balls.

Definition 3. The weight of a particle sea s is denoted $|s|$ and is the difference between the abscissæ of the squares of the positive quarter and the abscissæ of the balls of the non-positive quarter.

A *flat particle sea* is nothing but a sea particle with no ball or square above the zero-line. Itzykson's proof for Jacobi's triple product can be explained using only flat particle sea.

For a particle sea s, we define $d(s)$ (resp. $g(s)$) the number of squares (resp. balls) in the positive (resp. non-positive) quarter.

There exist a natural bijection between sea particles and pairs of overpartitions, one overpartition having positive parts: We split the sea particle into a positive and a non-positive side. For balls (resp. squares) in the non-positive (resp. positive) side we add their abscissa to their respective overpartition. If the ball (resp. square) in the non-positive (resp. positive) side is on the zero-line we overline its part.

4. q-Gauß

Let us use this new sea particle tool to prove the q-Gauß identity:

$$\sum_{n=0}^{\infty}\frac{(-1/b)_n(-1/a)_n(abq)^n}{(q)_n(q)_n} = \frac{(-aq)_\infty(-bq)_\infty}{(q)_\infty(abq)_\infty}. \tag{1}$$

Like for the q-binomial theorem, we identify each side of the identity as a product of generating functions of sets of overpartitions and sea particles. We are then left with providing a bijection between those sets. We note S_n instead of s for a sea particle s with $d(s) = g(s) = n$. The next algorithm allows us to go bijectively from the sea particle S_n to a pair of overpartitions (α, β) with additional conditions:

While there is a ball in the non-positive quarter do

- Choose the lowermost ball in the leftmost column, its coordinates are (x, y)

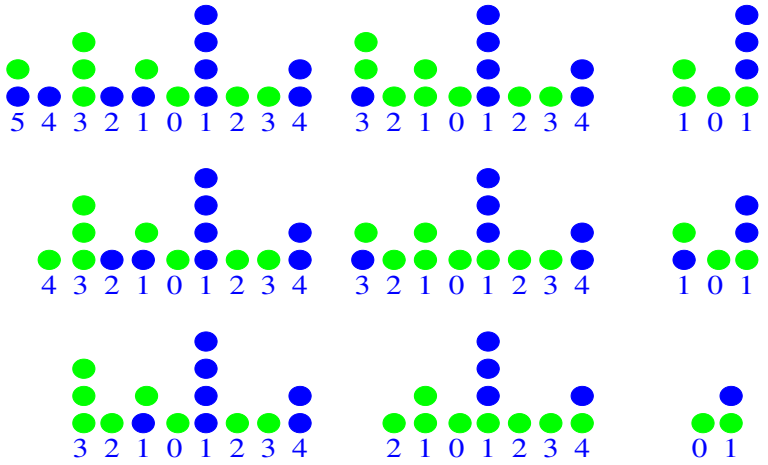


FIGURE 2. An example of the bijection for q-Gauß

- If there is a square on the zero-line of abscissa $x' = \min\{i, i > x\}$
  - $i := x' - x$
  - If $y = 0$, add a part i to α, change the square into a ball and the ball into a square
  - else add a part $\bar{i}$ to α, destroy the ball, change the square into a ball
- Else find the uppermost square in the rightmost column, its coordinates are (x', y')
  - $i := x' - x$
  - If $y = 0$, add a part $\bar{i}$ to β, destroy the square, change the ball into a square
  - else add a part i to β, destroy the ball and the square

The parameters of the overpartitions α and β are related to those on the original sea particle s by some conditions. The algorithm's steps are shown on picture 2 for some sea particle s and the final pair of overpartitions is $(\overline{7}, \overline{4}, 2, 2, \overline{1})$ and $(\overline{6}, \overline{2}, 2, \overline{1})$. It is easy to prove

$$\sum_{s \in S_n} q^{|s|} a^{g(s)-g_0(s)} b^{d(s)-d_0(s)} = \frac{(-1/b)_n (-1/a)_n}{(q)_n (q)_n} (abq)^n$$

and from there we use the bijection between sea particles from S_n and pairs of overpartitions with additional conditions to conclude.

5. Extensions

Sea particles have been used by the author to prove some more basic hypergeometric identities like Ramanujan's ${}_1\Psi_1$. The method is very similar the hard part being the search for the right bijection. The concept of particle sea has been extended to *colored particle sea* where, instead of having only balls and squares we can add some more colors to these. This leads to a more powerful tool and some harder results (related to ${}_6\phi_5$) have been obtained.

Bibliography

[1] Andrews (George E.). – *The Theory of Partitions*. – Addison–Wesley, 1976, *Encyclopedia of Mathematics and its Applications*, vol. 2.

[2] Corteel (Sylvie). – Particle seas and basic hypergeometric series. *Advances in Applied Mathematics*, vol. 31, 2003, pp. 199–214.

[3] Joichi (James T.) and Stanton (Dennis). – Bijective proofs of basic hypergeometric series identities. *Pacific Journal of Mathematics*, vol. 127, 1987, pp. 103–120.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 25–28.

Available online at the URL
http://algo.inria.fr/seminars/.

Counting Unrooted Maps Using Tree Decomposition

Éric Fusy
Algorithms Project, INRIA (France)

October 4, 2004

Summary by Frédéric Giroire

Abstract

This talk presents a new method to count unrooted maps on the sphere. This method is based on tree decomposition More precisely it gives enumerations of 2-connected and 3-connected unrooted maps with a complexity of $\mathcal{O}(N \log(N))$ for maps with $e \leq N$ edges and $\mathcal{O}(N^2)$ for maps with i vertices and j faces $i + j \leq N$. The family of 3-connected unrooted maps corresponds to the skeletons of polytopes in the 3D space, also called convex polyedra. This motivates us to find a good method to count these objects.

This work has been done with the help of Gilles Schaeffer.

1. Introduction

Maps and roots. A *map on the sphere* is the embedding of a graph on a sphere up to a continuous deformation. A *rooted map* is a map where one half-edge is marked.

Rooting greatly facilitates the enumeration by giving a starting point for a recursive decomposition. So, for example, we know for a long time the number of rooted planar maps:

$$\mathcal{M}'_n = \frac{2}{n+2} \frac{3^n (2n)!}{n!\,(n+1)!}$$

Classical method for enumeration of unrooted objects. Nevertheless some classical methods to count unrooted objects exist. One of them, introduced by Liskovets allows to obtain the number c_n of unrooted maps with n edges on the sphere by using Burnside's lemma (a result in group theory which is often useful in taking account of symmetry when counting mathematical objects) and the method of quotient. It gives c_n as a function of c'_n, the number of rooted maps, and $c_n^{(k)}$, the number of k-rooted maps (A *k-rooted map* is a map having $k \geq 2$ undistinguishable roots.):

$$c_n = \frac{1}{2n}(c'_n + \sum_{k \geq 2} \phi(k) c_n^{(k)})$$

This formula adapts also for other families of maps (for example 2-connected and 3-connected maps). For the case of unconstraint maps on the sphere, k-rooted maps are easily counted by noticing that the quotient of a k-rooted map with respect to the symmetry is a rooted map. The number of preimages of a rooted map for this surjection is easy to calculate. Hence we obtain from this method, called quotient method and due to Liskovets, the enumeration of k-rooted maps.

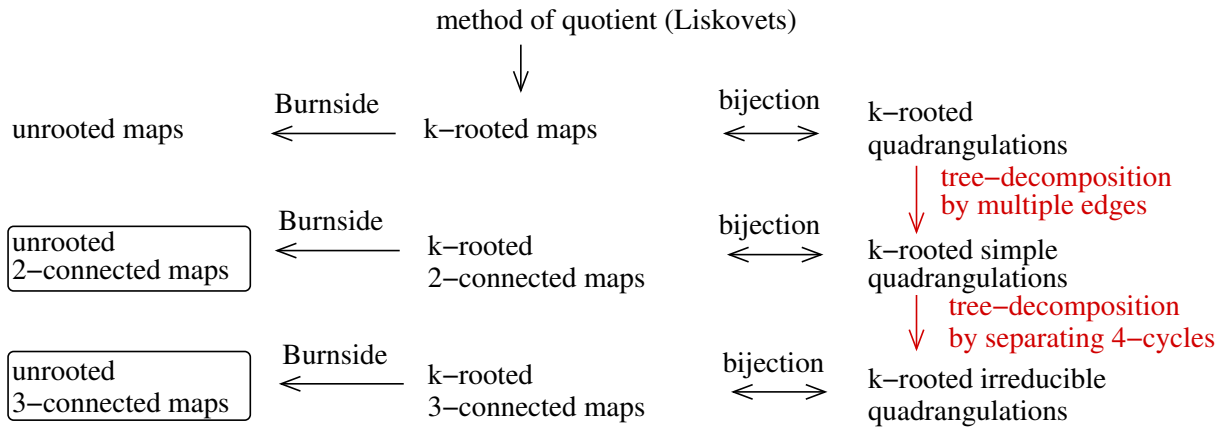


FIGURE 1. Scheme of the method.

Types of k-rooted maps. It was noticed by Liskovets that a k-rooted map has an embedding on the sphere wich is invariant by a certain rotation of angle $2\pi/k$ of the sphere. In addition, the two *poles* of the sphere crossed by the rotation-axis are either a vertex or the center of a face, and it can also be the middle of an edge if $k = 2$. These two points are called the *poles* of the k-rooted map. The *type* of the k-rooted map is the type of its two poles. For example, if the two poles are a vertex and a face, then the k-rooted map is said to have type face-vertex. In the particular case of a k-rooted quadrangulation, its type is more restricted. More precisely, the type is vertex-vertex if $k > 2$ and can also be face-vertex and face-face if $k = 2$.

Bijection between maps and quadrangulation. An other classical result that will be adapted to our problem is a wll-known bijection between maps and quadrangulations, which restricts well on 2-connected and 3-connected maps as illustrated on Figure 1.

Results and methods. Here will be only presented the method to count 2-connected unrooted maps. A similar one for 3-connected maps has also be done by the author. Figure 1 shows a summary of the results and methods used by the author. This new method of enumeration has a better complexity than previous ones. We will expose this point in our last section.

2. Enumeration of 2-Connected Unrooted Maps

One want here to count 2-connected unrooted maps. As in the classic case, using Burnside lemma, the enumeration comes down to count k-rooted 2-connected maps. Then one introduced a variation of the classical bijection between maps and quadrangulations: this bijection sets k-rooted 2-connected maps in bijection with k-rooted simple quadrangulations. We then use a tree decomposition method that will be explosed in detail in the following. We have now to deal with a family that we know how to enumerate (thanks to the bijection with the k-rooted maps and the method of the quotient).

Method to perform the tree-decomposition. We will transform an unrooted quadrangulation (that may have multiple edges) in a tree with two kinds of nodes: one representing the multiple edges and the other representing simple quadrangulations which are quadrangulations without multiple edges.

For each multiple edge of multiplicity d, we can imagine that we "blow," from the interior of the sphere, each of the d sectors delimited by the multiple edge. We so obtain d quadrangulations

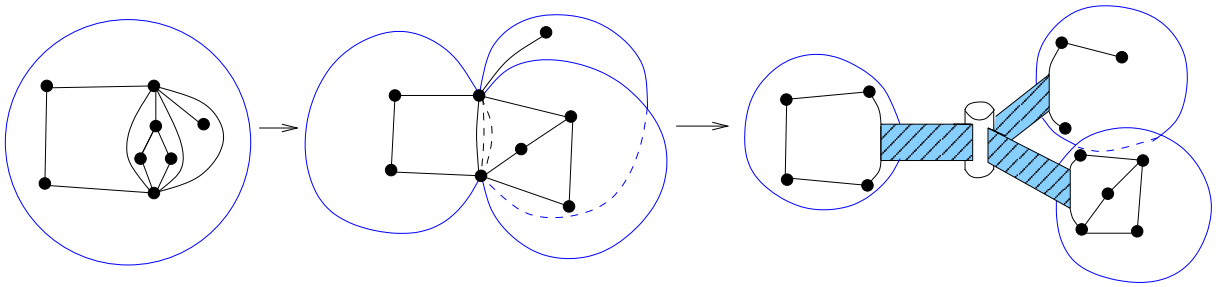


FIGURE 2. The tree-decomposition of a quadrangulation.

drawn on d spheres that are now d nodes of our tree. We connect them to a node representing the multiple edge. An example of this tree-decomposition can be seen in Figure 2. We then carry on the decomposition for each of the d components.

Equations. With this method we obtain equations linking the generating functions (GF) of k-rooted quadrangulations (known) and the GF of k-rooted simple quadrangulations (unknown) of the same type. Here we present equations for quadrangulations of type vertex-vertex:

$$F_{vv}^{(k)}(z) = zf'(z)\frac{1}{1-f(z)} + \frac{(z(1+F(z)))'}{F(z)+1} g_{vv}^{(k)}\left(z(1+F(z))\right),$$

where $F_{vv}^{(k)}(z)$ is the GF of k-rooted quadrangulations of type vertex-vertex and $g_{vv}^{(k)}$ the GF of k-rooted simple quadrangulations of type vertex-vertex All the GF of the equation are known except $g_{vv}^{(k)}\left(z(1+F(z))\right)$. We deduct from that $g_{vv}^{(k)}\left(z(1+F(z))\right)$ and $g_{vv}^{(k)}(y)$ by doing the change of variables $y = z(1+F(z))$. Similarly, all GF of k-rooted simple quadrangulations can be obtained. Then, application of Burnside lemma (see Figure 1) allows to obtain the coefficients c_n counting unrooted 2-connected maps:

n	2	3	4	5	6	7	8	9	10	11	12
c_n	2	2	3	6	16	42	151	596	2605	12098	59166

3. Algebraicity and Complexity

Here we prove that this enumeration of 2-connected and 3-connected maps can be done very quickly as enounced in Theorem 1. We first present here an efficient way to compute the GF of k-rooted simple quadrangulations. It is linear as opposed to the naive method in $\mathcal{O}(N^3)$. It is using the property of algebraicity of the GF of quadrangulations. Then we finish by showing that the global complexity of the theorem is implied by it.

Theorem 1. *In the case of the enumeration of 2-connected and 3-connected maps according to their number of edges, to obtain the first N coefficients, we need $\mathcal{O}(N\log(N))$ operations.*

In the case of the enumeration of 2-connected and 3-connected maps according to their number of faces and vertices, to obtain the table of the first coefficients with indexes (i,j) with $i+j \leq N$, we need $\mathcal{O}(N^2)$ operations.

Efficient Expansion of the GF of k-rooted simple quadrangulations. The starting point is the equation according to the number of faces:

$$F_{vv}^{(k)}(x) = xf'(x)\frac{1}{1-f(x)} + g_{vv}^{(k)}\left(x(1+F(x))^2\right)$$

As we know all the GF of the equation except $g_{vv}^{(k)}\left(x(1+F(x))^2\right)$, we can expand:

$$g_{vv}^{(k)}\left(x(1+F(x))^2\right) = 2x + 18x^2 + 180x^3 + \cdots .$$

We then do the change of variables: $y = x(1+F(x))^2$. We have $y = x + 4x^2 + 22x^3 + \cdots$ and $x(y) = y - 4y^2 - 10y^3 + \cdots$ and the equation becomes $g_{vv}^{(k)}(y) = 2y + 10y^2 + 56y^3 + \cdots$.

This naive method of computation has a complexity of $\mathcal{O}(n^3)$. We found a way to decrease it to linear. Instead of doing the change of variables directly between x and y, we use small algebraic series $\beta(x)$ and $\eta(y)$ associated respectively to x and y such that the relation between β and η is rational.

To do so we will use the algebraicity of the GF. Let call β the GF of blossoming trees. We have: $\beta(x) = x + 3\beta(x)^2$. The GF of rooted quadrangulations is a rational expression of $\beta(x)$:

$$F(x) = \frac{\beta(x)(2-9\beta(x))}{(1-3\beta(x))^2}.$$

All GF of k-rooted quadrangulations are also rational expressions of $\beta(x)$. So our starting equation becomes now:

$$g_{vv}^{(k)}\left(x(1+F(x))^2\right) = \frac{2\beta(x)}{1-6\beta(x)}.$$

The change of variables is $y = x(1+F)^2$, $y = \frac{\beta(1-4\beta)^2}{(1-3\beta)^3}$. We define η as $\eta = \frac{\beta}{1-3\beta}$. We have $y = \eta(1-\eta)^2$. η is an algebraic series in y (serie of trees): $\eta(y) = \frac{y}{(1-\eta(y))^2}$. Furthermore $\beta = \frac{\eta}{1+3\eta}$. We substitute $\frac{\eta}{1+3\eta}$ to β in $\frac{2\beta}{1-6\beta}$. We have at the end:

$$g_{vv}^{(k)}(y) = \frac{2\eta(y)}{1-3\eta(y)}.$$

From that we deduct a fast algorithm to compute the N first coefficients of the serie $g_{vv}(y)$:

1. take the resultant of $\begin{cases} -\eta(1-\eta)^2 + y = 0 \\ -g_{vv}(1-3\eta) + 2\eta = 0 \end{cases}$ We have $4g_{vv}^3 + 8g_{vv}^2 - 8y - 36yg_{vv} - 54yg_{vv}^2 + 4g_{vv} - 27yg_{vv}^3 = 0$;
2. find a differential equation verified by g_{vv} (We can use the function 'algeqtodiffeq' of 'gfun'): $g_{vv}(0) = 0, -4 - 6g_{vv}(y) + (2-54y)\frac{d}{dy}g_{vv}(y) + (-27y^2+4y)\frac{d^2}{dy^2}g_{vv}(y) = 0$;
3. take the coefficient $[y^n]$ in the equation to find a recursive equation for the coefficients (we can use the function 'diffeqtorec'): $(-6-27m-27m^2)u(m) + (6m+2+4m^2)u(m+1) = 0, u(0) = 0, u(1) = 2$.

Global complexity of the coefficients computation. The relation $2nc_n = c'_n + \sum_{k=2}^{n}\phi(k)c_n^{(k)}$ can be translated for the GF:

$$\sum_n 2nc_n y^n = g(y) + zg_{fv}(y^2) + z^2 g_{ff}(y^2) + \sum_{k=2}^{n}\phi(k)g_{vv}^{(k)}(y^k)$$

Let $\mathcal{C}_N(f)$ be the time of computation of the N first coefficients of a function $f(z)$. We have

$$\mathcal{C}_N\left(\sum_n 2nc_n\right) = \mathcal{C}_N(g) + \mathcal{C}_{N/2}(g_{fv}) + \mathcal{C}_{N/2}(g_{ff}) + \sum_{k=1}^{n}\mathcal{C}_{N/k}\left(g_{vv}^{(k)}\right)$$

We have $\mathcal{C}_N = \mathcal{O}(N)$ for the GF of k-rooted simple quadrangulations. $\mathcal{C}_N\left(\sum_n 2nc_n y^n\right) = \mathcal{O}(N) + \mathcal{O}(N/2) + \sum_{k=1}^{N}\mathcal{O}(N/k)$ So we need $\mathcal{O}(N\log(N))$ operations.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 29–43.

Available online at the URL
http://algo.inria.fr/seminars/.

Deux Approches pour l'Énumération des Cartes planaires[†]

Gilles Schaeffer

CNRS and Lix, École polytechnique (France)

January 18 and 19, 2004

Summary by Olivier Bernardi[‡]

Abstract

Nous allons traiter de l'énumération de certaines familles de cartes planaires. Cet exercice sera l'occasion d'illustrer deux approches concurrentes en combinatoire énumérative.

La première approche pour le comptage d'une famille d'objets est de nature bijective. Il s'agit d'exprimer les objets que nous cherchons à énumérer en fonction d'autres objets mieux connus. Dans le cadre de ce cours, nous montrerons que les quadrangulations sont en bijection avec les arbres planaires bien étiquetés.

La seconde approche, communément appelée méthode symbolique, est basée sur l'utilisation de séries génératrices. Elle consiste à rechercher une décomposition interne à notre famille d'objets et à traduire cette décomposition en une équation fonctionnelle vérifiée par la série génératrice. Certains jugeront cette méthode moins satisfaisante, arguant que le processus à l'œuvre n'apporte aucune compréhension sur la nature des objets énumérés. Néanmoins, elle permet un traitement systématique d'une large classe de problèmes. Le point faible réside plutôt dans la difficulté à résoudre les équations fonctionnelles obtenues. Ici nous introduirons, sur l'exemple du comptage des triangulations, une technique de résolution qui généralise la « méthode du noyau ».

1. Introduction : les Cartes planaires

Définissons tout d'abord la notion de carte planaire.

Definition 1 (Cartes planaires)**.** Une *carte planaire* est le plongement dans le plan d'un graphe planaire connexe fini non orienté et non vide. Le plongement doit respecter la non-intersection des arêtes.

Dans une carte, on reconnaîtra des sommets et des arêtes (image des sommets et des arêtes du graphe sous-jacent) et des faces qui sont les composantes connexes maximales du complémentaire de la carte. L'unique face non bornée est qualifiée d'*externe*, les autres sont dites *internes*.

Bien sûr, nous ne nous intéressons pas à la carte planaire elle-même, mais à la structure combinatoire sous-jacente. Nous considérerons donc comme identiques deux cartes qui peuvent s'obtenir, l'une à partir de l'autre, par une déformation continue du plan. Pour être plus formel, nous nous intéressons aux classes d'équivalence pour la relation d'homéomorphismes du plan orienté. Ainsi, sur la Figure 1 les deux cartes à gauche sont identiques mais différentes de la carte de droite. Notons au passage, qu'à un graphe peuvent correspondre plusieurs cartes différentes (ou aucune s'il est non planaire).

[†]Notes de cours pour le cours donné pendant le groupe de travail Aléa'04 au Cirm à Luminy (France).

[‡]LaBRI, Université Bordeaux 1, 33405 Talence Cedex, France ; email: bernardi@labri.fr.

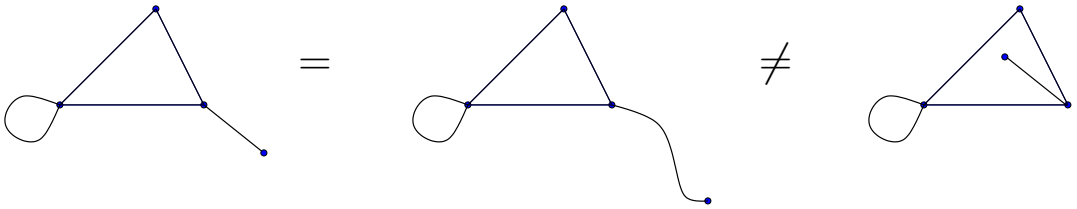


FIGURE 1. Équivalence entre cartes.

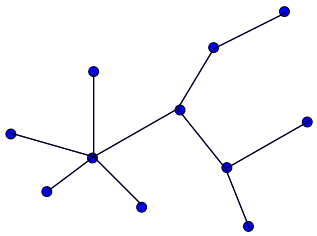


FIGURE 2. Un arbre plongé.

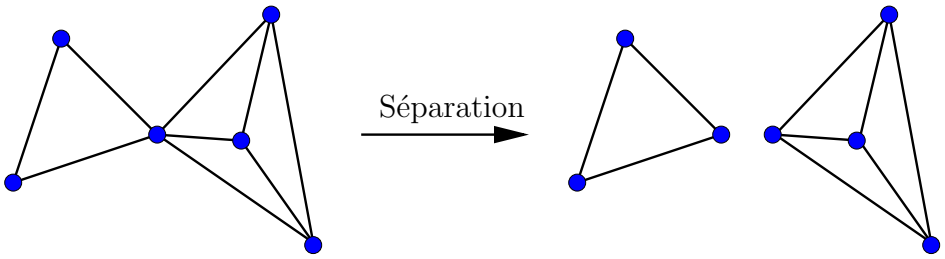


FIGURE 3. Une carte séparable.

Nous pouvons nous restreindre à des familles particulières de cartes en imposant certaines contraintes. Nous pouvons, par exemple, considérer l'ensemble des cartes sans face interne c'est-à-dire la famille des arbres plongés. Voir l'exemple en Figure 1. Alternativement, nous pouvons nous restreindre au cas des cartes *non-séparables*. Une carte est dite *séparable* si on peut la découper en deux parties non réduites à un point et dont l'intersection est réduite à un point. Un exemple de carte séparable est fourni par la Figure 3. Nous pouvons aussi étudier des cartes dont le degré des faces est fixé. Les cartes dont toute face est de degré 3 sont appelées *triangulations*. Les cartes dont toute face est de degré 4 sont appelées *quadrangulations*. La Figure 4 exhibe un représentant de chacune de ces familles. Nous reviendrons plus tard sur ces deux familles. Pour le moment, il nous faut définir la notion d'enracinement.

Definition 2 (Enracinement d'une carte)**.** Une carte planaire *enracinée* est une carte planaire dont une des arêtes de la face externe est orientée. L'orientation doit correspondre au sens direct pour le contour de la carte. L'arête qui est orientée est la *racine* de la carte et son sommet d'origine est le *sommet racine*. Graphiquement la racine sera repérée par une flèche indiquant le sens d'orientation.

Notons qu'une arête qui est un isthme peut être orientée dans les deux sens, tandis que pour une arête incidente à la face externe et à une face interne le sens d'orientation est imposé. Le nombre de façons d'orienter une carte planaire est égal au degré de sa face externe.

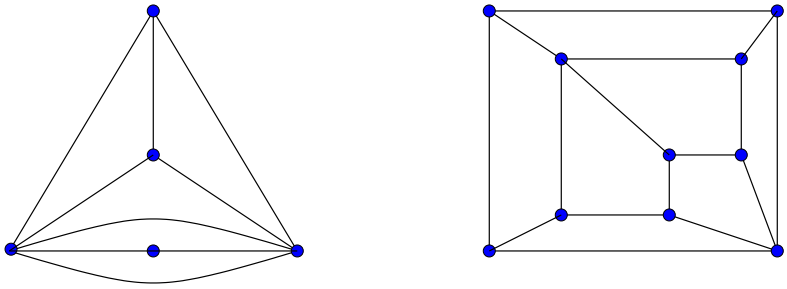


FIGURE 4. Une triangulation et une quadrangulation.

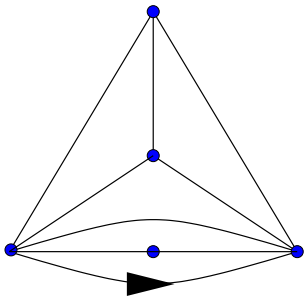


FIGURE 5. Une triangulation enracinée.

La notion d'enracinement est assez naturelle et le lecteur aura le plaisir de constater qu'il connaît le nombre de cartes planaires enracinées sans face interne à n arêtes. En effet, il s'agit des arbres planaires enracinés dont il est bien connu qu'ils sont comptés par les nombres de Catalan $\frac{1}{n+1}\binom{2n}{n}$.

Maintenant que nous sommes familiers avec les cartes planaires, nous aimerions les compter. Nous commençons par énumérer les triangulations grâce à la méthode symbolique (*i.e.* à base de séries génératrices) puis nous nous intéresserons aux quadrangulations.

2. **Comptage des Triangulations par la Méthode symbolique**

Dans cette section, nous cherchons à énumérer les triangulations non-séparables enracinées, c'est-à-dire les cartes planaires enracinées non-séparables dont toute face est de degré 3. Nous noterons $\mathcal{T}$ l'ensemble des triangulations non-séparables enracinées et nous considérerons le nombre d'arêtes comme le paramètre de taille sur $\mathcal{T}$. Nous cherchons donc à répondre à la question : combien y a-t-il de triangulations non-séparables enracinées à n arêtes ? Nous allons utiliser la méthode symbolique : on introduit la série génératrice

$$G(z) = \sum_{M \in \mathcal{T}} z^{|M|}$$

où $|M|$ est le nombre d'arêtes de la triangulation M, et on cherche à décomposer les triangulations de taille n en triangulations de taille inférieure pour obtenir une équation fonctionnelle vérifiée par la série génératrice $G(z)$. L'idée immédiate est alors de prendre une triangulation, de supprimer l'arête racine et de placer la nouvelle racine sur une autre arête. Cette opération est illustrée par la Figure 6.

Malheureusement, la carte obtenue n'est pas une triangulation puisque sa face externe est de degré quatre. Après avoir examiné la situation un moment, on se convainc qu'il est nécessaire

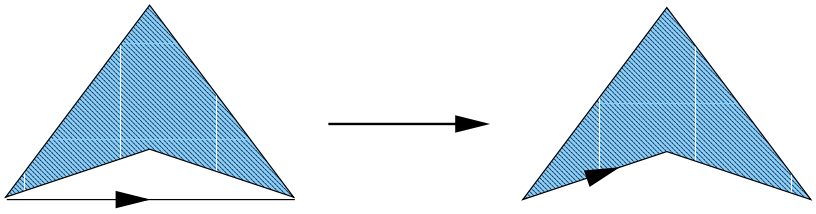


FIGURE 6. Décomposition d'une triangulation par la racine.

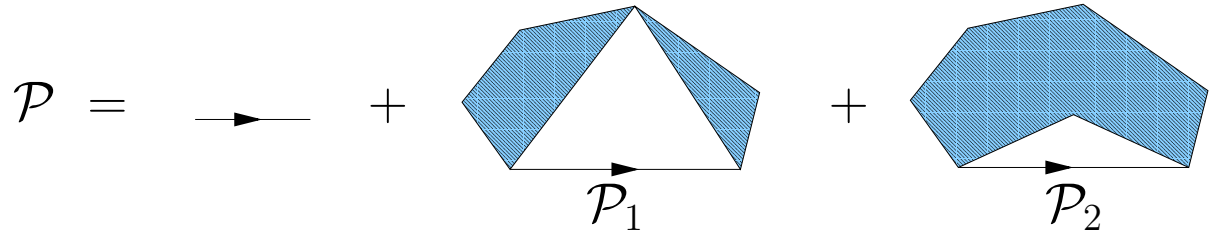


FIGURE 7. Décomposition d'une quasi-triangulation par la racine.

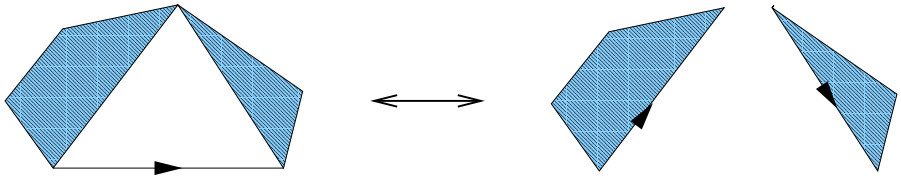


FIGURE 8. Bijection entre $\mathcal{P}_1$ et $\mathcal{P}^2$.

d'élargir notre famille de cartes à l'ensemble $\mathcal{P}$ des *quasi-triangulations*, soit les cartes non-séparables enracinées dont toute face interne est de degré 3 (dont on exclut la carte réduite à un sommet). De plus, nous aurons besoin d'introduire un paramètre supplémentaire, à savoir le degré de la face externe. Nous définissons donc la série génératrice bivariée, de variable principale z et de variable secondaire u,

$$H(u,z) = \sum_{M \in \mathcal{P}} u^{d(M)} z^{|M|}$$

où $|M|$ est le nombre d'arêtes de la quasi-triangulation M et $d(M)$ est le degré de la face externe de M. Reprenons maintenant notre décomposition. Une quasi-triangulation est soit réduite à une arête, soit comporte un triangle interne incident à la racine. Dans ce cas, le troisième sommet du triangle interne est soit sur la face externe soit pas. Cette décomposition est illustrée par la Figure 7.

Examinons les contributions respectives de ces trois cas à la série génératrice $H(u,z)$ des quasi-triangulations. La carte réduite à une arête a pour contribution $u^2 z$. On note $\mathcal{P}_1$ l'ensemble des quasi-triangulations, non réduites à une arête, dont le troisième sommet du triangle interne incident à la racine est sur la face externe. On exprime la contribution des cartes dans $\mathcal{P}_1$, en remarquant qu'il existe une bijection entre $\mathcal{P}_1$ et $\mathcal{P}^2$. Cette bijection est illustrée par la Figure 8. De plus, si on note (M_1, M_2) l'image d'une carte M par cette bijection, le nombre d'arêtes et le degré des faces externes des cartes M, M_1, M_2 sont liés par les relations $|M| = |M_1| + |M_2| + 1$ et

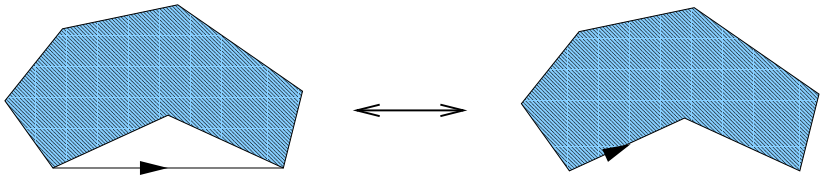


FIGURE 9. Bijection entre $\mathcal{P}_1$ et $\{M \in \mathcal{P}/d(M) > 2\}$.

$d(M) = d(M_1) + d(M_2) - 1$. On en déduit que la contribution des cartes dans $\mathcal{P}_1$ est

$$\sum_{M\in\mathcal{P}_1} u^{d(M)} z^{|M|} = \sum_{(M_1,M_2)\in\mathcal{P}^2} u^{d(M_1)+d(M_2)-1} z^{|M_1|+|M_2|+1} = \frac{z}{u}\left(\sum_{M\in\mathcal{P}} u^{d(M)} z^{|M|}\right) = \frac{z}{u} H(u,z)^2.$$

On note $\mathcal{P}_2$ l'ensemble des quasi-triangulations, non réduites à une arête, dont le troisième sommet du triangle interne incident à la racine n'est pas sur la face externe. On exprime la contribution des cartes dans $\mathcal{P}_2$, en remarquant qu'il existe une bijection entre $\mathcal{P}_2$ et l'ensemble des cartes dont le degré de la face externe est supérieure à 2. Cette bijection est illustrée par la Figure 9. De plus, si on note N l'image d'une carte M par cette bijection, le nombre d'arêtes et le degré des faces externes des cartes M, N sont liés par les relations $|M| = |N| + 1$ et $d(M) = d(N) - 1$. On en déduit que la contribution des cartes dans $\mathcal{P}_2$ est

$$\begin{aligned}\sum_{M\in\mathcal{P}_2} u^{d(M)} z^{|M|} &= \sum_{N\in\mathcal{P}/d(N)>2} u^{d(N)-1} z^{|N|+1} \\ &= \frac{z}{u}\left(\sum_{N\in\mathcal{P}} u^{d(N)} z^{|N|} - \sum_{N\in\mathcal{P}/d(N)=2} u^{d(N)} z^{|N|}\right) = \frac{z}{u}\left(H(u,z) - u^2 F(z)\right)\end{aligned}$$

où $F(z)$ est le coefficient de u^2 dans la série $H(u,z)$.

Au terme de cette analyse nous avons obtenu une équation fonctionnelle portant sur la série génératrice inconnue $H(u,z)$,

$$H(u,z) = u^2 z + \frac{z}{u} H(u,z)^2 + \frac{z}{u}\left(H(u,z) - u^2 F(z)\right) \tag{1}$$

où $F(z)$ est le coefficient de u^2 dans la série $H(u,z)$.

Arrêtons nous un moment pour contempler sereinement cette équation. C'est une équation qui fait apparaître 2 séries inconnues $H(u,z)$ et $F(z)$. Et il est intéressant de remarquer qu'elle contient l'information $F(z) = [u^2]H(u,z)$ si l'on a l'information préalable que $H(u,z)/u^2$ est une série en z à coefficients polynomiaux en u. Il suffit, pour l'obtenir, d'extraire le coefficient de u dans l'équation. De plus, elle définit bien les séries inconnues de manière unique comme séries formelles en la variable z. En effet, on constate que l'équation se met sous la forme

$$H(u,z) = u^2 z + z \cdot A(u,z)$$

où le coefficient de z^n dans $A(u,z)$ s'exprime en fonction des coefficients de z^k dans $H(u,z)$ avec $k \leq n$. Donc l'équation donne une relation de récurrence bien définie sur les coefficients de $H(u,z)$. Donc cette équation définit à elle seule deux séries inconnues (si on a l'information préalable que $H(u,z)/u^2$ est une série en z à coefficients polynomiaux en u). Nous aimerions donc bien la transformer en un système de deux équations. Si l'équation était linéaire en la série inconnue $H(u,z)$ cela se ferait assez naturellement en appliquant une technique connue sous le nom de

méthode du noyau (voir par exemple [3]). Ici, nous allons appliquer une extension de cette méthode en adoptant un formalisme dû à M. Bousquet-Mélou et A. Jehanne [2].

L'équation (1) est polynomiale en les séries inconnues. Et, si on note P le polynôme à quatre variables $P(H,F,U,Z) = U^3Z + ZH^2 + Z(H - U^2F) - UH$, l'équation (1) s'écrit

$$P(H(u,z),F(z),u,z) = 0. \tag{2}$$

Dérivons cette équation par rapport à u. On obtient

$$\frac{\partial H(u,z)}{\partial u} \cdot D_H P(H(u,z),F(z),u,z) + D_U P(H(u,z),F(z),u,z) = 0,$$

où $D_H P$ et $D_U P$ représentent les dérivées du polynôme P par rapport à H et à U respectivement.

On cherche maintenant s'il existe une série formelle $U(z)$ telle que[1]

$$D_H P(H(U(z),z),F(z),U(z),z) = 0. \tag{3}$$

Dans ce cas, on obtiendrait le système

$$\begin{aligned} P(H(U(z),z),F(z),U(z),z) &= 0 \\ D_H P(H(U(z),z),F(z),U(z),z) &= 0 \\ D_U P(H(U(z),z),F(z),U(z),z) &= 0 \end{aligned} \tag{4}$$

de trois équations, pour trois inconnues $H(U(z),z)$, $F(z)$ et $U(z)$. Et si nous parvenons à résoudre le système (4) nous connaîtrons $F(z)$, puis en reportant ce résultat dans l'équation de départ (1), nous connaîtrons $H(u,z)$. Notons bien que, pour que le système (4) soit valide, il suffit de savoir que $U(z)$ existe, non de le calculer. Dans notre exemple, l'équation (3) s'écrit

$$U(z) = z \cdot (1 + 2H(U(z),z)).$$

Il est facile de voir que cette relation se traduit en une relation de récurrence bien définie sur les coefficients de $U(z)$ et que par conséquent la série formelle $U(z)$ existe (et est unique). Le système d'équations (4) est donc valide et nous pouvons essayer de le résoudre. Par élimination, on trouve

$$U(z) = z\left(1 + 2U(z)^3\right) \qquad \text{et} \qquad F(z) = \frac{1}{2}\left(3U(z) - \frac{U(z)^2}{z}\right).$$

Ici, nous sommes dans une situation particulièrement favorable puisque notre équation pour la série $U(z)$ nous permet d'appliquer le théorème d'inversion de Lagrange (voir par exemple [1]). On rappelle que ce théorème stipule que si une série formelle $U(z)$ vérifie une équation du type $U(z) = z\Phi(U(z))$ alors le coefficient de z^n de la série $U(z)^k$ vaut

$$[z^n]U(z)^k = \frac{k}{n}[x^{n-k}]\Phi(x)^n.$$

Dans notre exemple, cela signifie que le coefficient de z^n dans $F(z)$ est égal à

$$\begin{aligned} [z^n]F(z) = \frac{1}{2}\left(3[z^n]U(z) - [z^{n+1}]U(z)^2\right) &= \left(\frac{3/2}{3p+1}[x^{n-1}](1+2x^3)^n - \frac{1}{3p+2}[x^{n-1}](1+2x^3)^{n+1}\right) \\ &= \frac{2^p}{(2p+1)(p+1)}\binom{3p}{p} \text{ si } n = 3p+1 \text{ et } 0 \text{ sinon.} \end{aligned}$$

Nous avons pratiquement terminé puisque nous allons voir que la série génératrice des triangulations s'exprime très simplement en fonction de la série $F(z)$. En effet, la série $F(z)$ a été définie

[1] On remarque que dans le cas d'une équation linéaire en la série inconnue $H(u,z)$, la dérivée de P par rapport à H, $D_H P$, est le coefficient multiplicatif de H dans l'équation. Donc, dans le cas linéaire, nous sommes en train de chercher une série $U(z)$ qui annule le *noyau*.

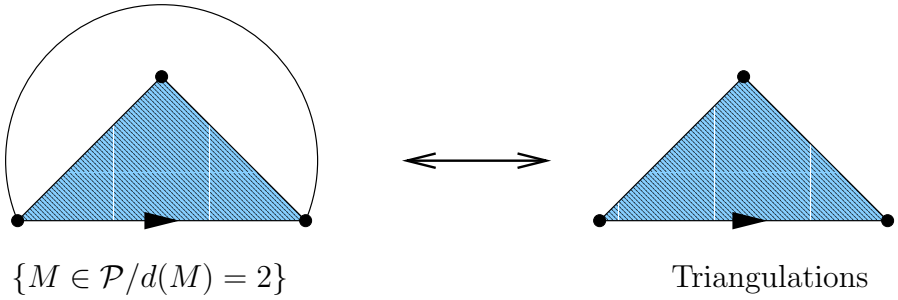


FIGURE 10. Bijection entre $\{M \in \mathcal{P}/d(M) = 2\}$ et l'ensemble $\mathcal{T}$ des triangulations.

comme le coefficient de u^2 dans la série $H(u, z)$ des quasi-triangulations. C'est donc la série formelle des quasi-triangulations dont la face externe est de degré 2, comptées selon le nombre d'arêtes. Il reste à remarquer que ces quasi-triangulations sont en bijection avec l'ensemble $\mathcal{T}$ des triangulations par l'opération illustrée sur la Figure 10.

La série génératrice $G(z)$ des triangulations est donc

$$G(z) = \frac{F(z)}{z} = \sum_{p>0} a_p z^{3p} \qquad \text{où} \quad a_p = \frac{2^p}{(2p+1)(p+1)} \binom{3p}{p}.$$

La méthode symbolique nous a donc permis de montrer que les triangulations ont un nombre d'arêtes multiple de 3 (ce qui s'explique facilement par la relation d'incidence faces/arêtes) et de donner le nombre a_p de triangulations à $3p$ arêtes.

La méthode que nous avons employée pour résoudre l'équation fonctionnelle semble pouvoir s'appliquer à un grand nombre d'exemples. Elle peut être formulée en des termes plus généraux afin de prendre en compte les cas où l'équation ferait apparaître plusieurs séries inconnues secondaires $F_1(z), F_2(z), \ldots, F_k(z)$. Ainsi, si nous avions une équation polynomiale du type

$$P(H(u, z), F_1(z), F_2(z), \ldots, F_k(z), u, z) = 0,$$

dont on peut montrer qu'elle définit toutes les séries inconnues $H(u, z)$, $F_1(z)$, ..., $F_k(z)$ de manière unique, nous serions amené à chercher k séries $U_1(z)$,..., $U_k(z)$ solutions de l'équation (en la variable u)

$$D_H P(H(u, z), F_1(z), F_2(z), \ldots, F_k(z), u, z) = 0.$$

Si de telles séries existent nous obtiendrons un système de $3k$ équations pour les $3k$ séries inconnues $U_1(z)$, ..., $U_k(z)$, $H(U_1(z), z)$, ..., $H(U_k(z), z)$, $F_1(z)$, , ..., $F_k(z)$ qu'il ne restera qu'à résoudre.

Nous passons maintenant à la présentation d'une bijection importante pour l'étude des cartes planaires. Il s'agit de la bijection entre quadrangulations et arbres bien étiquetés. Cette construction n'est pas immédiate et elle est d'autant plus intéressante qu'il existe une bijection simple entre les quadrangulations et les cartes planaires générales [8].

3. **Comptage des Quadrangulations par une Approche bijective**

Nous cherchons à énumérer les quadrangulations enracinées. Il nous faut un paramètre de taille adéquat. On remarque que si l'on note f le nombre de faces d'une quadrangulation alors le nombre d'arêtes est $a = 2f$ (par la relation d'incidence faces/arêtes qui s'écrit $4f = 2a$) et le nombre de sommets est $s = f + 2$ (par la relation d'Euler qui s'écrit $f + s = a + 2$). Nous prenons donc, comme paramètre de taille, le nombre de faces de la quadrangulation. Nous noterons $\mathcal{Q}_n$ l'ensemble des quadrangulations enracinées à n faces ($2n$ arêtes, $n + 2$ sommets). En utilisant la

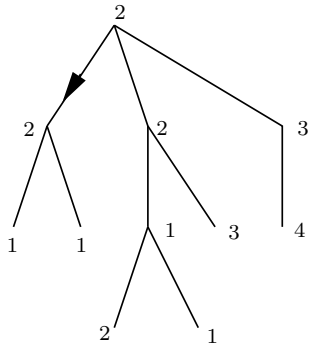


FIGURE 11. Un arbre étiqueté.

méthode symbolique que nous avons illustrée sur l'exemple des triangulations, on peut montrer que le nombre de quadrangulations à n faces ($2n$ arêtes, $n+2$ sommets) est

$$|\mathcal{Q}_n| \quad = \quad \frac{2}{n+2} \cdot \frac{3^n}{n+1} \binom{2n}{n}. \tag{5}$$

Cette formule peut nous apporter une intuition des objets combinatoires sous-jacents. En effet, on reconnaît dans le facteur $\frac{1}{n+1}\binom{2n}{n}$ les nombres de Catalan dont nous rappelons qu'ils comptent les arbres planaires enracinés à n arêtes. Donc $\frac{3^n}{n+1}\binom{2n}{n}$ est le nombre d'arbres planaires enracinées dont les arêtes sont étiquetées par $\{-1, 0, +1\}$. Ou, alternativement, on peut voir que ces nombres comptent les arbres planaires enracinés dont les sommets sont étiquetés dans l'ensemble des entiers naturels, dont la plus petite étiquette est égal à 1 et tels que les étiquettes de deux sommets adjacents diffèrent d'au plus 1. Un tel objet sera appelé *arbre étiqueté*. Un exemple d'arbre étiqueté est donné en Figure 11.

Nous noterons $\mathcal{A}_n$ l'ensemble des arbres étiquetés à n arêtes et nous retenons que

$$|\mathcal{A}_n| = \frac{3^n}{n+1} \binom{2n}{n}.$$

En outre, nous dirons qu'un arbre étiqueté est *bien-étiqueté* si l'étiquette de sa racine est 1. Nous allons maintenant montrer que les quadrangulations non-séparables enracinées sont en bijection avec les arbres bien-étiquetés. Ensuite nous démontrerons, grâce à cela, la formule énumérative (5). La bijection entre quadrangulations et arbres bien étiquetés a initialement été établie par R. Cori et B. Vauquelin [5]. Nous adopterons ici une présentation due à M. Marcus et G. Schaeffer [4, 6].

Soit M une quadrangulation non-séparable enracinée. On étiquette les sommets de la quadrangulation M selon leur distance au sommet racine. L'étiquetage a été effectué sur l'exemple de la Figure 12. Cet étiquetage canonique possède des propriétés intéressantes. On remarque tout d'abord que deux sommets adjacents ne peuvent avoir même étiquette. En effet, si deux sommets s et s' ont même étiquette i, on considère la région du plan délimitée par l'arête (s, s'), un chemin de longueur i du sommet racine à s et un chemin de longueur i du sommet racine à s'. L'intérieur de cette région est quadrangulé et l'extérieur a degré $2i+1$ qui est impair. Cette situation est impossible car la relation d'incidence faces/arêtes implique que l'extérieur d'une surface quadrangulée est toujours de degré pair. La différence entre les étiquettes de deux sommets adjacents est donc toujours égale à 1. Il s'ensuit que le graphe d'une quadrangulation est biparti (sommets pairs et sommets impairs). Et les faces sont toujours d'une des deux formes présentées sur la Figure 13.

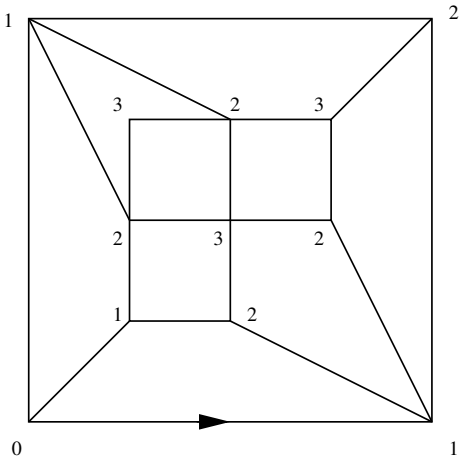


FIGURE 12. Distance au sommet racine dans une quadrangulation.

FIGURE 13. Deux types de faces dans une quadrangulation.

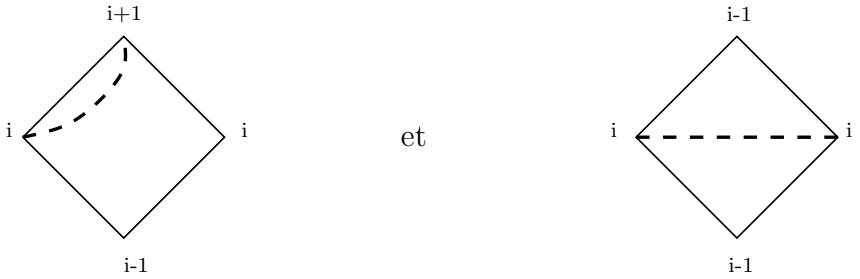


FIGURE 14. Règle de construction des arêtes.

Nous allons maintenant construire une arête par face de la quadrangulation et montrer que l'ensemble de ces arêtes forme un arbre bien-étiqueté. On construit une arête par face en appliquant la règle illustrée par la Figure 14. Sur l'exemple de la Figure 12, l'ensemble des arêtes ainsi construites est représenté en traits discontinus sur la Figure 15.

Soit $A(M)$ l'ensemble des arêtes ainsi construites sur une quadrangulation M. Nous allons montrer que cet ensemble constitue un arbre. Vérifions d'abord qu'il est sans cycle. On suppose que $A(M)$ contient un cycle. On choisit un cycle simple et on considère une plus petite étiquette i sur ce cycle. Par une étude de cas, illustrée par la Figure 16, on montre qu'il existe un sommet

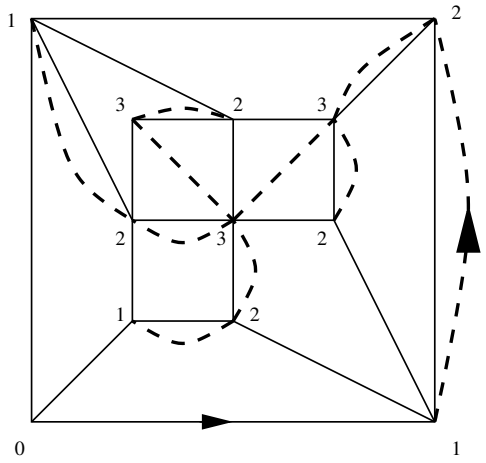


FIGURE 15. Construction sur l'exemple.

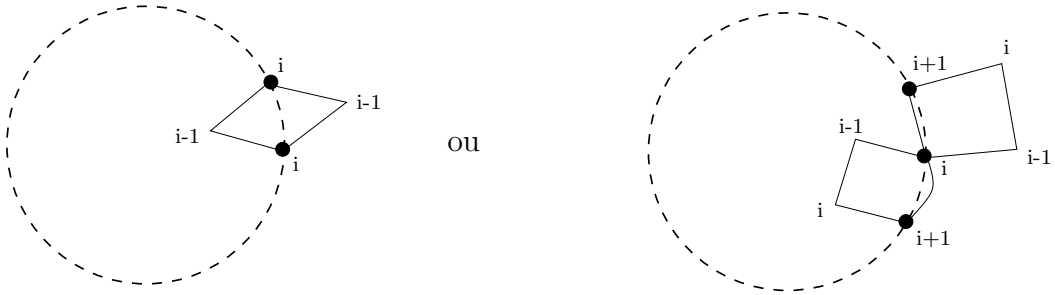


FIGURE 16. Étude de cas : $A(M)$ ne contient pas de cycle.

d'étiquette $i-1$ à l'intérieur et aussi à l'extérieur du cycle. Cela impliquerait que la racine est à la fois à l'intérieur et à l'extérieur du cycle, ce qui est impossible.

L'ensemble des arêtes $A(M)$ est sans cycle. C'est donc une forêt. De plus, les arêtes de $A(M)$ ne sont pas incidentes au sommet racine. Elles sont donc incidentes au maximum à $n+1$ sommets. Enfin, le cardinal de l'ensemble $A(M)$ est égal à n (le nombre de faces de la quadrangulation). Nous en déduisons que $A(M)$ est un arbre et que cet arbre couvre tous les sommets non-racines.

La racine de la quadrangulation M pointe sur l'arbre $A(M)$ en un sommet étiqueté 1. Ce pointage permet d'attribuer canoniquement une racine à $A(M)$. C'est ce qui a été fait sur la Figure 15. Il est facile de voir que, de cette façon, l'arbre $A(M)$ est bien étiqueté.

Pour montrer que ce processus est bijectif, nous allons exhiber la construction réciproque. Soit T un arbre bien étiqueté avec $n+1$ sommets. On considère la *liste circulaire des sommets* c'est-à-dire la liste des sommets (ou plutôt des incidences des sommets sur la face externe, aussi appelées *coins*), obtenue en parcourant le contour de l'arbre dans le sens direct. Pour l'arbre bien étiqueté de la Figure 17, la liste circulaire des sommets est $(a, b, a, c, d, c, e, c, a, f, g, f)$. Pour chaque apparition d'un sommet étiqueté par $i > 1$ dans la liste on crée une arête qui relie ce sommet au premier sommet étiqueté $i-1$ qui le suit dans la liste circulaire. Cette opération est illustrée par la Figure 18. Enfin, on crée un sommet étiqueté 0 et on relie à ce sommet tous les sommets étiquetés 1 dans la liste circulaire. On s'impose, à cette étape, que l'arête reliant le sommet 0 au sommet racine (l'incidence de ce sommet racine « au dessus de l'arbre ») soit sur la face externe et soit orientée dans le sens

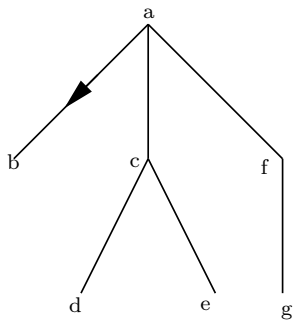


FIGURE 17. Un arbre bien étiqueté.

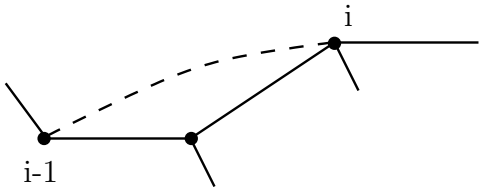


FIGURE 18. Création d'une arête entre un sommet i et un sommet $i-1$.

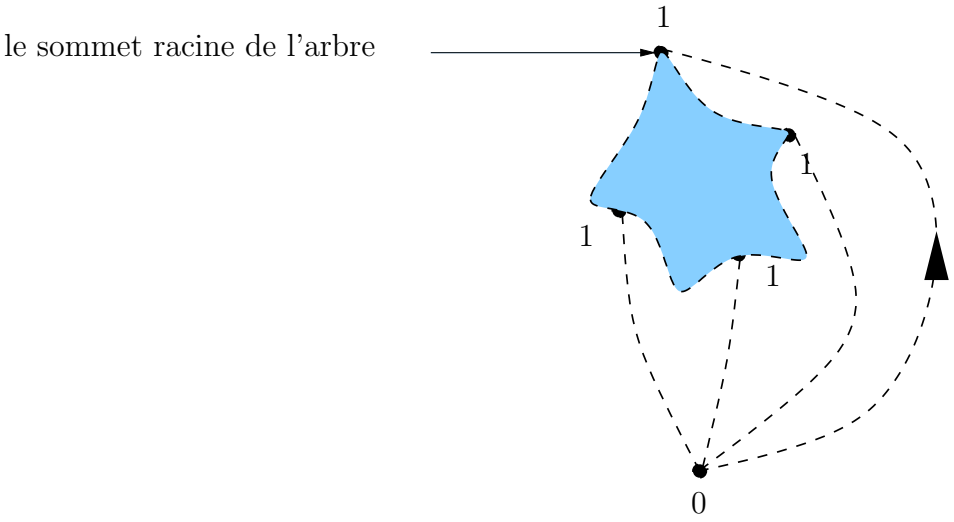
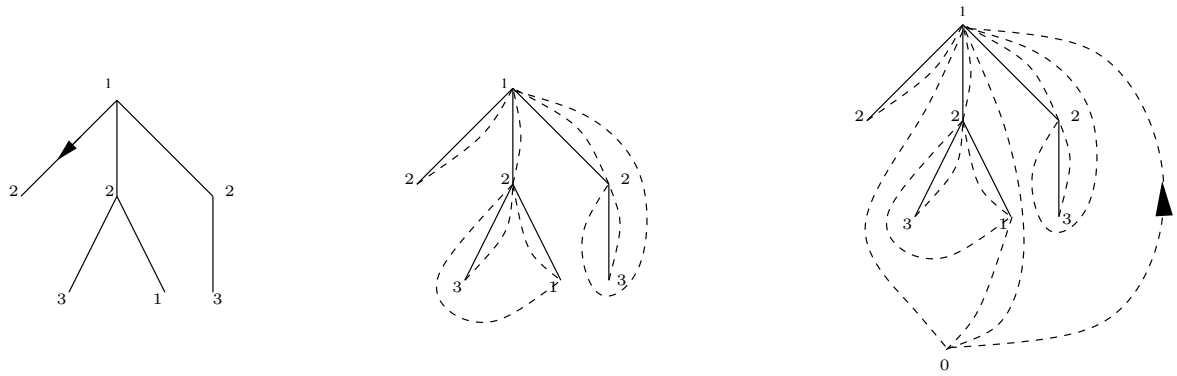


FIGURE 19. Création des arêtes entre les sommets d'étiquette 1 et le sommet 0.

positif du plan. On attribue la racine à cette arête orientée. Cette opération est illustrée par la Figure 19. La construction complète a été réalisée sur l'exemple de la Figure 20.

Montrons à présent que l'ensemble des arêtes ainsi créées forme bien une quadrangulation. Il nous faut d'abord vérifier que c'est une carte planaire, c'est-à-dire que les arêtes créées ne s'intersectent pas. Considérons une arête créée d'un sommet s d'étiquette i à un sommet s' d'étiquette $i-1$ et montrons que cette arête n'entre en conflit avec aucune autre arête. Soit s'' un sommet situé entre s et s' dans la liste circulaire. On remarque que les étiquettes d'un sommet et de son voisin dans la liste circulaire diffèrent d'au plus 1. Par conséquent, l'étiquette de s'' est $j \geq i$ (car sinon il existe



Un arbre bien étiqueté. On relie les sommets $i > 1$ aux sommets $i - 1$. On relie les sommets 1 au sommet 0.

FIGURE 20. Construction d'une quadrangulation à partir d'un arbre bien étiqueté.

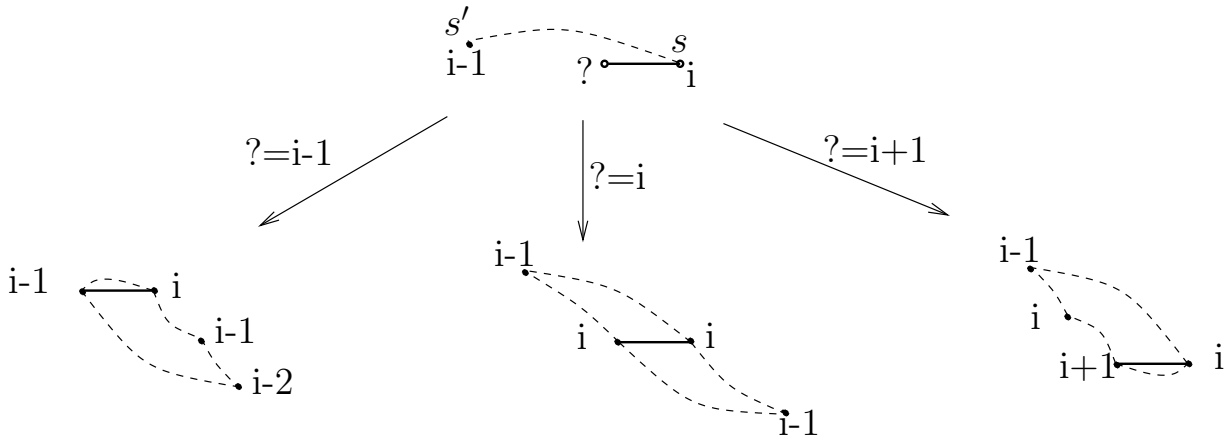


FIGURE 21. Étude de cas : les faces sont de degré 4.

un sommet d'étiquette $i-1$ situé strictement entre s et s') et il existe un sommet d'étiquette $j-1$ entre s'' et s' (puisque $i-1 \leq j-1$). De ce fait, l'arête créée à partir de s'' n'entrera pas en conflit avec l'arête (s, s'). Les arêtes ne s'intersectent donc pas et on obtient une carte planaire.

Il nous faut maintenant vérifier que les faces de notre carte sont de degré 4. On considère une face. Il existe au moins une arête orientée (pour l'orientation directe de la face) d'un sommet s étiqueté i à un sommet s' étiqueté $i-1$. On distingue 3 cas selon que l'étiquette du sommet suivant s dans la liste circulaire vaut $i-1$, i ou $i+1$. Dans les trois cas, une analyse simple quoique fastidieuse permet de montrer que la face est de degré 4. Ces trois cas sont représentés sur la Figure 21.

Nous allons esquisser le raisonnement dans le premier cas. Nous considèrons une face f de la carte et les sommets s et s' de cette face tels que définis plus haut. Nous supposons que l'étiquette du sommet suivant s dans la liste circulaire des sommets est $i-1$. Dans ce cas, ce sommet est s'. La situation est représentée sur la Figure 22. Sur cette figure les parties grisées sont les sous arbres de part et d'autre de l'arete (s, s') (de l'arbre étiqueté) et on s'intéresse aux incidences (sur la face externe de l'arbre) des sommets s et s'. On distingue l'incidence (ou « coin ») a de s et l'incidence a'

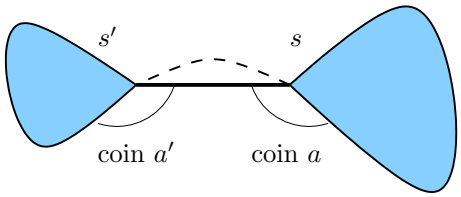


FIGURE 22. Les incidences ou « coins » a et a' des sommets s et s'.

de s'. On remarque que l'arête construite du coin a' de s' à un sommet s'' d'étiquette $i-2$ appartient nécessairement à notre face f. On constate aussi qu'aucune arête n'arrivera au coin a (car l'étiquette de s' est plus petite que celle de s). Il s'ensuit que l'arête partant du coin a vers un sommet s''' d'étiquette $i-1$ appartient aussi à notre face f. Reste à remarquer qu'une arête sera construite du sommet s''' au sommet s'' et que cette arête referme la face f. Notre face f est donc bien de degré 4. Un raisonnement analogue s'applique dans les autres cas et permet de prouver que toute face est de degré 4.

Il reste à montrer que cette construction est bien la réciproque de la précédente. On note $Q(T)$ la quadrangulation obtenue à partir de l'arbre bien étiqueté T. On remarque que les étiquettes sur $Q(T)$ sont les distances des sommets au sommet 0. Cette propriété est facile à vérifier par induction puisque tout sommet d'étiquette i est relié à au moins un sommet d'étiquette $i-1$ et à aucun sommet d'étiquette inférieure. De plus, en revenant sur les différents cas de faces possible dans $Q(T)$ (Figure 21) il apparaît que chaque face de $Q(T)$ contient exactement une arête. Et que cette arête est précisément celle que nous construirions si nous appliquions l'algorithme de construction d'un arbre bien étiqueté à partir de la quadrangulation $Q(T)$. De même, il est facile de voir que la racine que nous attribuerions en construisant un arbre à partir de $Q(T)$ est bien la racine de T. Il apparaît donc que pour tout arbre T on a bien la relation $A(Q(T)) = T$.

Nous avons donc démontré que pour tout arbre bien étiqueté $A(Q(T)) = T$. Il resterait à montrer que pour toute quadrangulation M on a $Q(A(M)) = M$. Nous ne le ferons pas ici. Nous invitons le lecteur curieux à se référer à la thèse de G. Schaeffer [7].

Nous allons maintenant exposer une variante de la bijection ci-dessus et en déduire la formule énumérative des quadrangulations. On cherche le cardinal de $\mathcal{Q}_n$, l'ensemble des quadrangulations enracinées à n faces. Soit $\mathcal{Q}'_n$ l'ensemble des quadrangulations enracinées à n faces dont un sommet est distingué. Nous savons qu'une quadrangulation à n faces possède $n+2$ sommets. Donc les cardinaux de $\mathcal{Q}_n$ et $\mathcal{Q}'_n$ sont liés par la relation $|\mathcal{Q}'_n| = (n+2)|\mathcal{Q}_n|$. Soit M une quadrangulation dont un sommet est distingué. On construit un arbre $A'(M)$ à partir de la quadrangulation M comme suit. On étiquette les sommets de M selon leur distance au sommet distingué (et non plus par leur distance au sommet racine). Puis on crée une arête par face en utilisant la même règle que précédemment (Figure 14). Enfin, on place la racine de l'arbre $A'(M)$ sur l'arête créée sur la face externe. Cette construction a été réalisée sur l'exemple de la Figure 23.

Ensuite on fixe une orientation à la racine. Il nous faut en plus garder en mémoire un paramètre dans $\{0,1\}$ pour que le processus soit bijectif. La règle d'attribution de l'orientation ainsi que du paramètre binaire est décrite par la Figure 24. Dans l'exemple précédent, le paramètre binaire est égal à 1.

On obtient un élément de l'ensemble $\mathcal{A}_n$ des arbres étiquetés plus un élément de $\{0,1\}$. On montre que cette construction est une bijection de $\mathcal{Q}'_n$ dans $\mathcal{A}_n \times \{0,1\}$ et par conséquent

$$|\mathcal{Q}'_n| = |\mathcal{A}_n \times \{0,1\}| = 2 \cdot \frac{3^n}{n+1}\binom{2n}{n}.$$

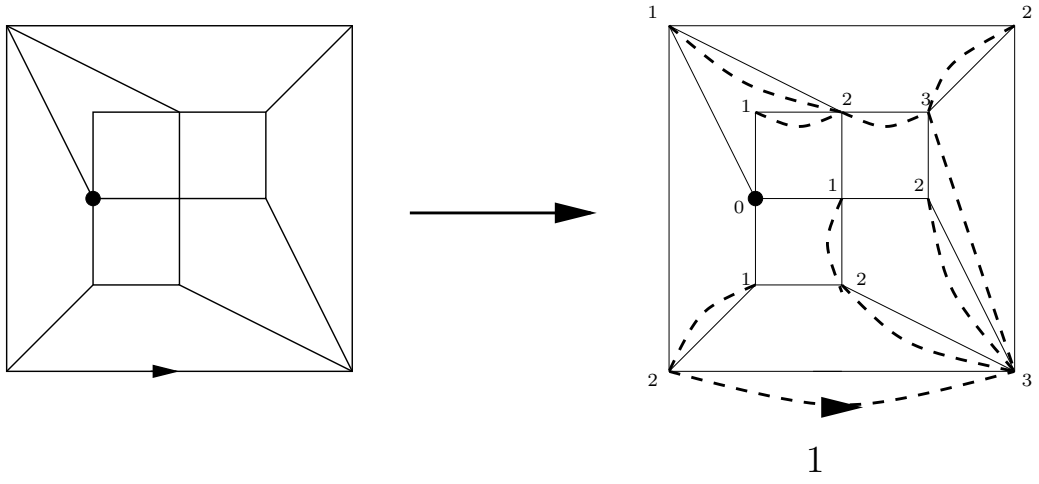


FIGURE 23. Construction d'un arbre à partir d'une carte pointée.

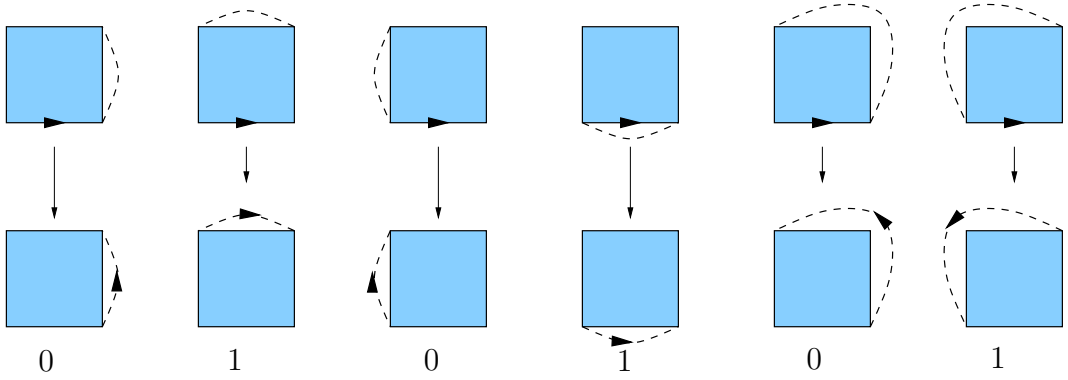


FIGURE 24. Règle d'attribution de l'orientation de la racine et du paramètre binaire.

D'où on déduit la formule énumérative annoncée pour les quadrangulations à n faces :

$$|\mathcal{Q}_n| = \frac{2}{n+2} \cdot \frac{3^n}{n+1} \binom{2n}{n}.$$

4. **Conclusions**

Nous avons illustré deux approches concurrentes, bijective d'une part et symbolique d'autre part, pour l'énumération de certaines familles de cartes planaires. Les difficultés ne sont pas de même nature dans les deux méthodes et l'une peut échouer là où l'autre réussit. Cependant, elles peuvent aussi se révéler complémentaires comme dans le cas du comptage des quadrangulations où la formule énumérative obtenue par la méthode symbolique permet d'obtenir une intuition d'une bijection non triviale.

Bibliographie

[1] *Combinatorial Enumeration.* – John Wiley and Sons, 1983.

[2] Bousquet-Mélou (M.) and Jehanne (A.). – Planar maps and algebraic series: Polynomial equations with one catalytic variable. – En préparation.
[3] Bousquet-Mélou (M.) and sek (M. Petkov$\check{}$). – Linear recurrences with constant coefficients: the multivariate case. *Discrete Mathematics*, vol. 225, 2000, pp. 51–75.
[4] Chassaing (P.) and Schaeffer (G.). – Random planar lattices and integrated superBrownian excursion. *Probability Theory and Related Fields*, vol. 128, n° 2, 2004, pp. 161–212.
[5] Cori (R.) and Vauquelin (B.). – Planar maps are well labeled trees. *Canadian Journal of Mathematics*, vol. 33, n° 5, 1981, pp. 1023–1042.
[6] Marcus (M.) and Schaeffer (G.). – Une bijection simple pour les cartes orientables. – Manuscrit en français, `http://www.loria.fr/~schaeffe/`.
[7] Schaeffer (G.). – *Conjugaison d'arbres et cartes combinatoires aléatoires*. – PhD thesis, Université Bordeaux I, Bordeaux, 1998.
[8] Tutte (W. T.). – A census of planar maps. *Canadian Journal of Mathematics*, vol. 15, 1963, pp. 249–271.

Part II

Analytic Combinatorics and Asymptotics

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 47–50.

Available online at the URL
http://algo.inria.fr/seminars/.

On the Asymptotic Analysis of a Class of Linear Recurrences

Thomas Prellberg
Technische Universität Clausthal (Germany)

September 23, 2002

Summary by Marni Mishna

Abstract

This work offers an interesting application of analytic iteration theory and classical complex analysis to determine some new (and old) results in asymptotic enumeration. The method treats functional equations of a particular form, which have a natural interpretation in terms of combinatorial generating functions. Partition lattice chains and Takeuchi numbers are among the applications of this method presented here.

1. Problems Suited to this Analysis

Many combinatorial classes can be described in a recursive way, built from basic atomic units using a handful of combinatorial operations, as described in [2]. One of the principal fruits of this point of view is a set of functional equations for the exponential and ordinary generating functions of the family. The work presented here considers families satisfying a particular type of combinatorial equation and gives explicit asymptotic formulas, determined directly from the corresponding functional equations. The principal results are summarized in Theorem 1. Two applications of the main theorem are detailed: Asymptotic enumeration of partition lattice chains and Takeuchi numbers. This technique is equally amenable to the asymptotic enumeration of Bell numbers.

We begin with brief descriptions of these two problems. For each example we give a functional equation satisfied by a generating function of the family.

1.1. **Partition lattice chains.** The set of partitions of an n-set can be ordered by subset inclusion to build a poset. Define Z_n as the number of chains from the minimal element $\{\{1\},\{2\},\dots,\{n\}\}$ to the maximal element $\{1,2,\dots,n\}$. This sequence begins $Z_1=1, Z_2=2, Z_3=4, Z_4=32$. These numbers satisfy the following recurrence, due to Lengyel [4]:

$$Z_n = \sum_{k=1}^{n-1} S_{n,k} Z_k$$

where the $S_{n,k}$ are the Stirling numbers of the second kind. From this, we deduce the functional equation for the exponential generating function $Z(z)=\sum_n Z_n \frac{z^n}{n!}$, also due to Lengyel:

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

In the final section we give an asymptotic formula for Z_n, which matches previous work by Flajolet and Salvy.

1.2. **Takeuchi Numbers.** Consider the following recursive function of Takeuchi, related to ballot numbers:

$$\mathrm{TAK}(x,y,z) := \textbf{ if } x \le y \textbf{ then } y \textbf{ else } \mathrm{TAK}(\mathrm{TAK}(x-1,y,z),\mathrm{TAK}(y-1,z,x),\mathrm{TAK}(z-1,x,y)).$$

Denote by $T(x,y,z)$ number of times the **else** clause is invoked when evaluating $\mathrm{TAK}(x,y,z)$. Define the sequence T_n by $T_n = T(n,0,n+1)$. The initial terms are $T_1 = 1, T_2 = 4, T_3 = 14, T_4 = 53$.

Knuth determined the following recurrence [3], and its corresponding functional equation for the ordinary generating function $T(z) = \sum_n T_n z^n$:

$$T_{n+1} = \sum_{k=0}^{n} \left[\binom{n+k}{n} - \binom{n+k}{n+1} \right] T_{n-k} + \sum_{k=1}^{n+1} \binom{2k}{k} \frac{1}{k+1};$$

$$T(z) = zC(z)\,T(zC(z)) + \frac{C(z)-1}{1-z}, \quad C(z) = \sum_{k=0}^{\infty} \binom{2k}{k} \frac{z^k}{k+1}. \tag{2}$$

The methodology presented here yields a new result for the asymptotic expansion of T_n.

1.3. **General setup.** The common feature of these two problems is that they satisfy a linear recurrence of the form

$$X_n = \sum_{k=1}^{n} c_{n,k} X_{n-k} + b_n,$$

with a functional equation for either the ordinary or exponential generating function $X(z)$ of the form:

$$X(z) = a(z) X \circ f(z) + b(z),$$

where $f(z) = z + cz^2 + dz^3 + \ldots$ has a parabolic fixed point. This is the functional equation associated with the following combinatorial equation where $\circ$ denotes the substitution operation: $\mathcal{X} = \mathcal{A} \times \mathcal{X} \circ \mathcal{F} + \mathcal{B}$. The remainder of this work is devoted to determining an asymptotic expression for X_n.

2. Asymptotic Analysis

The asymptotic analysis X_n has three major steps. First, we determine an expression for X_n as an integral, and then we perform a two step analysis on this integral, first using analytic iteration theory and then using a saddle-point analysis.

2.1. **An expression for the coefficient.** If a formal power series satisfies Eq. (1.3), with $a(z), f(z)$, and $b(z)$ analytic near $z = 0$, then we have the formal solution

$$X(z) = \sum_{m=0}^{\infty} \left(\prod_{k=0}^{m-1} a \circ f^k(z) \right) b \circ f^m(z).$$

We use this formal solution and the Cauchy inversion formula to determine an expression for the coefficients of the generating series. We have that $X_n = \sum_{m=0}^{\infty} X_{n,m}$ with

$$X_{n,m} = \frac{1}{2\pi i} \oint \left(\prod_{k=0}^{m-1} a \circ f^k(z) \right) b \circ f^m(z) \, \frac{dz}{z^{n+1}}. \tag{3}$$

2.2. **Analytic iteration theory.** To illustrate the general idea, we consider a slightly simpler problem. Let $Y(z)$ be a solution of the homogeneous equation

$$Y(z) = a(z)\, Y \circ f(z). \tag{4}$$

In this case we have

$$\prod_{k=0}^{m-1} a \circ f^k(z) = \frac{Y(z)}{Y \circ f^m(z)}.$$

With this, Eq. (3) rewrites as

$$X_{n,m} = \frac{1}{2\pi i} \oint \frac{Y(z)}{Y \circ f^m(z)} b \circ f^m(z)\, \frac{dz}{z^{n+1}} = \frac{1}{2\pi i} \oint \frac{b \circ f^m(z)}{Y \circ f^m(z)} Y(z)\, \frac{dz}{z^{n+1}}. \tag{5}$$

To establish the existence of $Y(z)$ and certain analyticity properties, we use analytic iteration theory, see [1, 5], and some astute observations.

First, we use the parabolic linearization theorem to show the conjugacy of $f(z)$ to a shift. We have that $f^{-1}(z)$ exists in some cardioid domain and maps contractively to it, via some (determinable) function μ. We deduce that $f^k(z) = \mu\left(\mu^{-1}(z) - k\right)$ for z sufficiently small. Given a complete asymptotic expansion for μ, we have that $f^{-m} \circ \mu(s) = \mu(s+m)$ admits a complete asymptotic expansion for $m \to \infty$ of the form:

$$\mu(s+m) \sim \frac{1}{cm} \left(1 + \left(1 - \frac{d}{c^2} - s \right) \frac{\log m}{m} + \sum_{k=2}^{\infty} \sum_{j=0}^{k} \nu_{j,k}(s) \frac{(\log m)^j}{m^k} \right). \tag{6}$$

Substitute $z = \mu(s)$ into Eq (4), and capitalize on the resulting similarities to the gamma function to determine a solution to the homogeneous equation. Most importantly, we deduce the following asymptotic result:

$$\frac{Y \circ \mu(s+n)}{Y \circ \mu(n)} \sim (a \circ \mu(n))^s. \tag{7}$$

Next, substitute $z = \mu(s+m)$ and Eq. (7) into the last integral in Eq. (5) and then apply the asymptotic expansion of $\mu(s+m)$ from Eq. (6) to get the following asymptotic formula:

$$X_{n,m} \sim (cm)^n m^{-1-(1-\frac{d}{c^2})\frac{n}{m}} \frac{Y \circ \mu(m)}{2\pi i} \int_{\mathcal{C}} \frac{b \circ \mu(s)}{Y \circ \mu(s)} \left(a \circ \mu(m) e^{\frac{n}{m}} \right)^s ds.$$

Returning to X_n, we see that the sum simplifies to

$$X_n \sim C \sum_m (cm)^n \frac{Y \circ \mu(m)}{m} \left(a \circ \mu(m)\right)^{(1-\frac{d}{c^2})\log m}$$

with

$$C = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{b \circ \mu(s)}{Y \circ \mu(s)}\, ds. \tag{8}$$

2.3. **Saddle-point analysis.** We conclude by applying a saddle-point analysis to the integral in Eq. (8). There is a saddle point at $a \circ \mu(m) e^{\frac{n}{m}} = 1$. The saddle has different behavior depending on the smallest term of $a(z) = a_k z^k + \cdots$. We can summarize this analysis in the following theorem, which treats the $k = 0, 1$ cases separately.

Theorem 1. *Suppose the formal power series* $X(z) = \sum_{n=0}^{\infty} X_n z^n$ *satisfies*

$$X(z) = a(z)X \circ f(z) + b(z)$$

with $f(z) = z + cz^2 + dz^3 + \ldots$, $a(z) = a_k z^k + \ldots$, *and* $b(z)$ *analytic near zero. If* $c > 0$ *and* $0 < a_k$ *then the following are true:*

1. *If* $k = 0$, *and* $a_0 < 1$, *then*

$$X_n \sim Dn!\,(-c/\log a_0)^n\, n^{(1-\frac{d}{c^2})\log a_0 - 1} \qquad \text{as } n \to \infty, \text{ where} \qquad D = (-\log a_0)^{-(1-\frac{d}{c^2})\log a_0};$$

2. *If* $k = 1$ *then*

$$X_n \sim Dc^n e^{-\frac{1}{2}(1-\frac{d}{c^2})W(\frac{c}{a_1}n)^2} \sum_{m=0}^{\infty} \frac{m^n}{m!} \left(\frac{a_1}{c}\right)^m \qquad \text{as } n \to \infty, \text{ where} \qquad D = e^{\frac{1}{2}(1-\frac{d}{c^2})(\log\frac{a_1}{c})^2}.$$

3. Combinatorial Applications

We now possess sufficiently many tools to determine some asymptotic results with our earlier examples.

3.1. **Partition lattice chains.** Using Eq. (1) we deduce $a(z) = \frac{1}{2}$, $f(z) = e^z - 1$, $b(z) = \frac{z}{2}$. We have $\mu(s) \sim \frac{2}{s}(1 - \frac{\log s}{3s} + \ldots)$, and $Y \circ \mu(s) = 2^s$ thus we insert $c = \frac{1}{2}$, $d = \frac{1}{6}$, $a_0 = \frac{1}{2}$ into the main theorem, part 1. The resulting asymptotic expansion is

$$Z_n \sim D(n!)^2\,(2\log 2)^{-n}\, n^{-1-\frac{1}{3}\log 2}$$

as $n \to \infty$, where

$$D = \frac{1}{2}\,(\log 2)^{\frac{1}{3}\log 2}\,\frac{1}{2\pi i}\int_{\mathcal{C}} 2^s \mu(s)\,ds = 1.0986858055\ldots.$$

3.2. **Takeuchi numbers.** From Eq. (2), we have $a(z) = zC(z)$, $f(z) = zC(z)$, and $b(z) = \frac{C(z)-1}{1-z}$. From this we determine $\mu(s) \sim \frac{1}{s}(1 - \frac{\log s}{s} + \ldots)$, and thus, $Y \circ \mu(s) \sim e^{-\frac{1}{2}(\log s)^2}/\Gamma(s)$. Denote by B_n the nth Bell numbers. Applying these values to part 2 of the main theorem yields the asymptotic expansion:

$$T_n \sim D \sum_{m=0}^{\infty} \frac{m^n}{m!} e^{\frac{1}{2}W(n)^2} = D' B_n e^{\frac{1}{2}W(n)^2}$$

as $n \to \infty$, where

$$D' = \frac{e}{2\pi i}\int_{\mathcal{C}} \frac{b \circ \mu(s)}{Y \circ \mu(s)}\,ds = 2.2394331040\ldots.$$

Bibliography

[1] Beardon (A.). – *Iteration of rational functions.* – Springer, 1991, *Graduate Texts in Mathematics*, vol. 132.

[2] Flajolet (Philippe) and Sedgewick (Robert). – *Analytic Combinatorics—Symbolic Combinatorics.* – Research Report n° 4103, INRIA, 2002. 186+vii pages.

[3] Knuth (D. E.). – Textbook examples of recursion. In *Artificial intelligence and mathematical theory of computation*, pp. 207–229. – Academic Press, Boston, MA, 1991.

[4] Lengyel (T.). – On a recurrence involving Stirling numbers. *European Journal of Combinatorics*, vol. 5, 1984, pp. 313–321.

[5] Milnor (J.). – *Dynamics in one complex variable, introductory lectures.* – Vieweg Verlag, 2000.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 51–54.

Available online at the URL
http://algo.inria.fr/seminars/.

Patterns in Trees

Thomas Klausner
Technical Mathematics, Technische Universität Wien (Austria)

December 9, 2002

Summary by Marianne Durand and Julien Clément

Abstract

Given a tree, considered as a pattern, the question is how many times this pattern appears in a tree of size n. The average and variance of this parameter are obtained in this talk.

1. Introduction

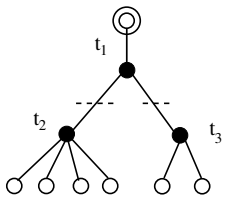
The problem of counting the number of occurrences of a pattern in a general tree is motivated for example by compression of arithmetical expressions. This talk presents first a simpler problem, that is counting planted patterns in planted trees, and reduces this problem to solving asymptotically a system of functional equations satisfied by related generating functions. The second part shows that the problem of general trees and general patterns is in fact very close to the planted problem, and can be reduced in a similar way to solving certain systems of functional equations. In the last section, asymptotic results on the number of occurrences of a given pattern in trees of size n are found from those systems, namely a normal distribution with explicit mean and variance.

2. Planted Trees and Planted Patterns

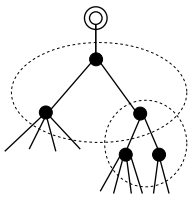
A planted tree is a rooted tree, where the degree of the root is equal to 1. To begin with, it is simpler to search planted patterns in planted trees.

2.1. **Combinatorial decomposition.** The search of a planted pattern in a planted tree is as follows. First, see if the pattern and the tree match when you match the two roots, this may give an occurrence, second build planted subtrees, and search recursively inside. Building planted subtrees consists in erasing the root (this gives a tree, as the root was of degree 1) and then split the new root into a root for each of his sons, to create a forest of planted trees.

FIGURE 1. Example of a pattern.



(a) A white circle stands for any tree.



(b) A tree in which the pattern occurs twice.

The pattern shown in Figure 1 is the example we use all along this summary. The pattern is first decomposed in planted subtrees (also named sub-patterns) which are named t_i (the ordering does not matter). Formally to obtain a planted subtree, one has to cut an internal edge in 2, and add a planted root on the cut side of the edge . To get all the subtrees, do it for all internal edges. The pattern is then fully known by the relation between its sub-patterns. For the example, we have the relations:

$$t_1 = o \times t_2 \times t_3$$
$$t_2 = o \times p \times p \times p \times p$$
$$t_3 = o \times p \times p$$

where o stands for the (planted) root, and p for any tree. In a planted tree the root is used to indicate where is the "top," so when a planted tree is seen as a subtree of a planted tree, this information is no longer necessary. This explains why the relation $t_1 = o \times t_2 \times t_3$ holds, with the subtrees t_2 and t_3 "losing" their planted root.

Now with this description, we are able to search recursively a planted pattern in a planted tree, but to count them, we have to take care of overlaps as shown in Figure 1(b). The reason appears during the writing of the generating function equation. Overlaps are possible, because of the non trivial intersection of the definition of t_1 and t_3. To avoid this, it is sufficient to rewrite the t_i differently to obtain a disjoint set of specifications (that is no tree satisfy two specifications). The t_i define sub-sets with overlaps of the set p of all trees, the symbols a_i are defined as standing for the underlying partition of the set, based on theoretic set operations (intersections, union, difference) involving the t_i's or the set of all trees p. Now all the a_i are disjoint (they are defined as a partition), and each t_i can be written as a union of a_i's.

The system of equations obtained on the t_i's is then easily translated into a system in the a_i's. In the example we obtain:

$$a_1 = t1 = \{o\} \times a_2 \times (a_1 \cup a_3)$$
$$a_2 = \{o\} \times p \times p \times p \times p$$
$$a_3 = (\{o\} \times (p \times p)) \setminus a_1$$
$$a_4 = p \setminus (a_1 \cup a_2 \cup a_3)$$

This is a system involving only the a_i's, as the relation $p = a_1 \cup a_2 \cup a_3 \cup a_4$ holds. In the a_i's basis, only a_1 represents a pattern, and so can be counted as a pattern, that is marked with a u symbol in the generating function, as explain in the next paragraph. Whereas in the t_i basis, t_3 may be a pattern, but without certainty, so that we do not know whether it should be counted or not.

2.2. **Generating functions.** The generating function of all trees is denoted by $p(z, u)$, where z codes the size of the tree, and u the number of patterns. So that $p(z, u) = \sum_{n,k} p_{n,k} \frac{z^n}{n!} u^k$, with $p_{n,k}$ the number of trees of size n (the number of nodes of the planted rooted tree) that contains k occurrences of the pattern. In what follows,if a letter a stands for a set of trees, then $a(z, u)$ stands for the corresponding generating function. As the set of all the trees p is decomposed as the disjoint union of the a_i's, the generating function $p(z, u)$ is the sum of the $a_i(z, u)$. For a presentation of the relation between combinatorial decomposition and generating functions, see [2]. Basically, the operations on sets are translated into operations on generating functions; unions of disjoint sets, exclusions and products translate into $+$, $-$ and $\times$. The system of equations between the a_i is translated into generating functions equations. The number of a_i's is denoted by L. All the relations but the last are written as equalities between a_j and the root times disjoint union or exclusion of a_i's, so that we have the relation $a_j(z, u) = zP_j(a_1(z, u), \ldots, a_L(z, u), u)$,

where P_j is a polynomial. The last variable of the polynomial, u, is used to mark the patterns. For the last relation, $a_L = p \setminus \cup a_i$, we get $a_L(z,u) = ze^{p(z,u)} - z\sum_{j=1}^{L-1} P_j(a_1(z,u),\ldots,a_L(z,u),1)$. To understand this equation, remember that the generating function of all trees, when there is no pattern to be counted, is $p(z) = ze^{p(z)}$. The rest is a basic translation, except for the last variable u. The P_j's have to be applied to 1 for their last variable, because in the set a_L, there is no pattern to be marked.

On the example, we have the system:

$$\begin{aligned}
a_1(z,u) &= uza_2(z,u)(a_1(z,u) + a_3(z,u)) \\
a_2(z,u) &= zp(z,u)^4 \\
a_3(z,u) &= z(p(z,u)^2 - a_1(z,u)) \\
a_4(z,u) &= ze^{p(z,u)} - z(a_2(z,u)(a_1(z,u) + a_3(z,u)) + p(z,u)^4 + p(z,u)^2 - a_1(z,u)).
\end{aligned}$$

The only pattern marked is in the first equation.

In this section, we have found how to obtain a system of equations satisfied by the generating function $p(z,u)$, that counts the number of occurrences of a given planted rooted pattern in planted rooted trees.

3. General Trees and General Patterns

Before searching general patterns in general trees, we consider the problem of searching two planted patterns in a planted tree. The number of patterns counted is the sum of the number of occurrences of the two patterns, eventually with overlaps. The idea is to make a "union" counting on the patterns. The technique is very similar to what is presented in section 2, so we just give the main lines.

The first pattern is decomposed into sub-patterns, named $t_1,\ldots,t_k$, the second pattern is also decomposed into the sub-patterns $t_{k+1},\ldots,t_j$. Then, all the t_i are grouped as if they came from the same pattern, the partition a_i is found from all the t_i's. The system of equation in the generating function $a_i(z,u)$ is built in a similar way, and both patterns are marked with a u. At the end we have a system of equation satisfied by the generating function $p(z,u)$ we are looking for.

Now that we know how to count (count is used in the sense of having an equation satisfied by the generating function) for two patterns, the next step is to count the occurrences of a not-planted pattern in a planted rooted tree. In order to do this, we build all possible ways of planting the patterns, and we consider the union of all these planted patterns as explained in the previous paragraph.

Finally to search a pattern (non-planted) in a tree (non-planted), we simply plant the tree, and then search the pattern inside the planted tree, as the planting of a tree does not change the number of occurrences of the pattern.

So the problem of counting the number of occurrences of a pattern in a tree is reduced to "solving" a system of equations of corresponding generating functions. Next section is devoted to finding asymptotic results on the coefficients of the generating function $p(z,u)$.

4. Asymptotic

To study the asymptotic behavior of the $p_{n,k}$, defined as coefficients of $p(z,u) = \sum p_{n,k}\frac{z^n}{n!}u^k$, we rely on a general theorem of Michael Drmota [1], that studies solutions of systems of functional

equations. Let X_n denotes the random variable of law defined by

$$\mathbf{P}(X_n = k) = \frac{p_{n,k}}{p_n} \qquad \text{with } p_n = \sum_k p_{n,k}.$$

This theorem, under some conditions, proves that X_n is asymptotically Gaussian, and provides explicit formulas for the mean and variance, that are here proportional to n.

To start with, we define some notations. Bold letters always stand for a vector. The letter $\mathbf{a}$ denotes the vector $(a_1(z,u),\dots,a_L(z,u))$. The symbol $\mathbf{F}$ stands for

$$\mathbf{F}(u,\mathbf{y},z) = (F_1(u,\mathbf{y},z),\dots,F_L(u,\mathbf{y},z))$$

where

$$F_i(u,\mathbf{y},z) = zP_i(\mathbf{y},u) \quad (\text{for } 1 \le i < L), \qquad \text{and} \qquad F_L(u,\mathbf{y},z) = ze^{\sum_{i=1}^{L} y_i} - z\sum_{i=1}^{L-1} P_i(\mathbf{y},1).$$

So that the system of functional equations can be rewritten as $\mathbf{a} = \mathbf{F}(u,\mathbf{a},z)$. The differentiation of a function f with respect to x is written f_x, and $\mathbf{F}_\mathbf{y}$ is the matrix with entries $\frac{\partial F_i}{\partial y_j}$.

We want the system to be strongly connected, which means that there is no sub-system that can be solved independently. This is verified when the pattern does not contain leaves, as each a_i but the last depends on the last (as each sub-pattern ends into at least one unspecified tree), and a_L depends on all a_i but itself. For the particular case where the pattern requires the presence of leaves, the branch leading to the leaf is considered as a black box, and plays no role in the decomposition. This particular case can thus be reduced to the general case without leaves.

A slightly modified version of the main theorem from [1] now says that under conditions that are satisfied in this context, X_n is asymptotically Gaussian, with mean and variance

$$\mathbf{E}[X_n] = \mu n + O(1) \qquad \text{and} \qquad \mathbf{Var}[X_n] = \sigma^2 n + O(1).$$

The value of the constants μ and σ comes from the solution of the system of functional equations

$$\begin{aligned} \mathbf{y} &= \mathbf{F}(u,\mathbf{y},z) \\ 0 &= \det(\mathbf{I} - \mathbf{F}_\mathbf{y}(u,\mathbf{y},z)) \end{aligned}$$

that admits a solution $\mathbf{y}(z)$ and $u(z)$. The constants μ and σ^2 are given by

$$\mu = -\frac{u_z(1)}{u(1)} \qquad \text{and} \qquad \sigma^2 = -\frac{u_{zz}(1)}{u(1)} + \mu^2 + \mu.$$

Bibliography

[1] Drmota (Michael). – Systems of functional equations. *Random Structures and Algorithms*, vol. 10, n° 1–2, 1997, pp. 103–124.

[2] Sedgewick (Robert) and Flajolet (Philippe). – Analytic combinatorics—Symbolic combinatorics. – 2005. `http://algo.inria.fr/flajolet/Publications/books.html`.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 55–60.

Available online at the URL
http://algo.inria.fr/seminars/.

Analytic Urns

Philippe Flajolet
Algorithms Project, Inria (France)

January 20, 2003

Summary by Pierre Nicodème

Abstract

The talk[1] describes an analytic approach to urn models of the Pólya type where an urn may contain balls of either of two colours. At each step, a ball is randomly drawn and replaced by balls of the two colours; a fixed 2×2-matrix with constant row sum determines the replacement policy. The treatment starts from a partial differential equation associated with the model and bases itself on conformal mapping arguments coupled with singularity analysis techniques. This gives access to moments characterizations, Gaussian limits and large deviation results. In some specific and well-determined cases, the urn models admit explicit representations in terms of Weierstraß elliptic functions.

1. Introduction

We follow in this summary the following route:

1. analyze the urns system

$$\mathcal{T}_{2,3} = \begin{pmatrix} -2 & 3 \\ 4 & -3 \end{pmatrix}, \tag{1}$$

   that models also the 2-3 trees, and leads to explicit results involving elliptic functions;
2. briefly sketch the analysis done by Flajolet et al. in the general case.

For complete proofs and further references, see the article "Analytic urns" of Philippe Flajolet, Joaquim Gabarró, and Helmut Pekari [1].

2. Analysis of the $\mathcal{T}_{2,3}$ Model

In the $\mathcal{T}_{2,3}$ model of Equation 1, when drawing a black ball one pulls out 2 black balls and adds 3 white balls, while when drawing a white ball one pulls out 3 white balls and adds 4 black balls; in the equivalent $\mathcal{T}_{2,3}$ tree model, black (resp. white) balls count the number of keys in 2-nodes (resp. 3-nodes), and the system models the transformation of a 2-node into a 3-node and the splitting of a 3-node into two 2-nodes. The initial counts are 2 black balls and 0 white balls, that insures a *tenable* system that cannot be blocked by removing more balls of any color than present in the urn.

[1] This talk presents a joint work of Philippe Flajolet with Joaquim Gabarró, and Helmut Pekari of University of Barcelona.

2.1. **The basic PDE.** Let X_n be the random variable representing the number of black balls at time n. The dynamics of the process is expressed by the stochastic recurrence

$$X_1 = 2; \qquad X_n - X_{n-1} = \begin{cases} -2 & \text{with probability } \dfrac{X_{n-1}}{n} \\ +4 & \text{with probability } 1 - \dfrac{X_{n-1}}{n}. \end{cases} \tag{2}$$

Let $p_{n,k} = \mathbf{P}(X_n = k)$. We also define

$$p_n(u) := \sum_k p_{n,k} u^k = \mathbf{E}(u^{X_n}) \quad \text{and} \quad F(z,u) := \sum_{n \geq 1} p_n(u) z^n = \sum_{n,k} p_{n,k} u^k z^n.$$

where $p_n(u)$ is the *probability generating function* (PGF) of X_n.

The recurrence 2 translates into a recurrence on the $p_{n,k}$, and the PGF $p_n(u)$ satisfies the differential recurrence,

$$p_n(u) = u^4 p_{n-1}(u) + \frac{1}{nu}(1-u^6)\frac{d}{du} p_{n-1}(u), \tag{3}$$

for $n \geq 2$, together with the initial condition $p_1(u) = u^2$. From this, we get a *partial differential equation* (PDE) satisfied by the bivariate generating function (BGF) $F(z,u)$:

$$(u^5 z - u)\frac{\partial F}{\partial z} + (1-u^6)\frac{\partial F}{\partial u} + u^5 F + u^3 = 0. \tag{4}$$

A slightly modified version of this is

$$G(z,u) := p_0(u) + F(z,u), \quad \text{with} \quad p_0(u) = (1-u^6)^{1/6} \int_0^u t^3 (1-t^6)^{-7/6}\, dt \tag{5}$$

given by solving Equation 3 for $n = 1$. The function G is now a solution of the homogeneous equation

$$(u^5 z - u)\frac{\partial G}{\partial z} + (1-u^6)\frac{\partial G}{\partial u} + u^5 G = 0. \tag{6}$$

2.2. **The solution by quadrature.** The algorithm given in Figure 1 provides a general way to solve quasi-linear partial differential equations of the first order. Applying this method to the PDE (6), we aim at solving the ordinary differential system:

$$\frac{du}{1-u^6} = \frac{dz}{u^5 z - u} = -\frac{dw}{u^5 w}. \tag{14}$$

the solutions $w = w(u)$ and $z = z(u)$ respectively verifies

$$w(1-u^6)^{-1/6} = C_1 = U(z,u,w) \quad \text{and} \quad z(1-u^6)^{1/6} + \int_0^u \frac{t}{(1-t^6)^{5/6}}\, dt = C_2 = V(z,u,w). \tag{15}$$

for some arbitrary integration constant C_1 and C_2. (Use the variation-of-constant method to compute $z(u)$). With

$$I(u) := \int_0^u \frac{t}{(1-t^6)^{5/6}}\, dt, \qquad \delta(u) := (1-u^6)^{1/6}.$$

we obtain the functional equation (with the notations of Figure 1 and $w = G$)

$$\Phi\left(\frac{G}{\delta(u)}, z\delta(u) + I(u)\right) = 0,$$

where Φ is an arbitrary function.

Start with a *quasi-linear* (bivariate) partial differential equation of the form

$$A(z,u,G)\frac{\partial G(z,u)}{\partial z} + B(z,u,G)\frac{\partial G(z,u)}{\partial u} + C(z,u,G) = 0, \tag{7}$$

where A, B, C are given functions.

1. First look for a solution in implicit form $X(z,u,G) = 0$. A calculation shows that the trivariate X must satisfy the *linear* (trivariate) partial differential equation:

$$A(z,u,w)\frac{\partial X(z,u,w)}{\partial z} + B(z,u,w)\frac{\partial X(z,u,w)}{\partial u} - C(z,u,w)\frac{\partial X(z,u,w)}{\partial w} = 0. \tag{8}$$

2. Next consider the ordinary differential system

$$\frac{dz}{A} = \frac{du}{B} = -\frac{dw}{C}. \tag{9}$$

The solution of two "independent" ordinary differential equations induced by (9), e.g.,

$$\frac{du}{B} = -\frac{dw}{C} \quad \text{and} \quad \frac{dz}{A} = \frac{du}{B}, \tag{10}$$

leads to two families of integral curves known as "first integrals,"

$$U(z,u,w) = C_1 \quad \text{and} \quad V(z,u,w) = C_2, \tag{11}$$

with z and w respectively treated as parameters. Assuming nondegeneracy, the generic solution of the PDE (8) is provided by

$$X(z,u,w) = \Phi(U(z,u,w), V(z,u,w)), \tag{12}$$

for an arbitrary bivariate Φ.

3. The trivariate X determines G implicitly by $X(z,u,G) = 0$, that is, by (12) one must have $\Phi(U(z,u,G), V(z,u,G)) = 0$. Solving for G provides a relation $G = R_\Phi(z,u)$, where R_Φ depends upon the arbitrary function Φ. The general solution of (7) is then

$$G(z,u) := R_\Phi(z,u). \tag{13}$$

FIGURE 1. The solution algorithm for quasilinear PDEs of first order.

Solving for G gives

$$G(z,u) = \delta(u)\Psi(\delta(u)z + I(u)), \tag{16}$$

where Ψ is an *arbitrary* function.

The unknown function Ψ is identified from the boundary condition,

$$G(0,u) = p_0(u), \tag{17}$$

that implies, assuming that (16) remains valid in this boundary case:

$$\frac{p_0(u)}{\delta(u)} = \Psi(I(u)) \quad \text{or} \quad J(u) := \int_0^u t^3(1-t^6)^{-7/6}\,dt = \Psi(I(u)). \tag{18}$$

The relation (18) then provides a *parameterization* of Ψ, hence it eventually determines a plausible value for G. This gives (with the previous notations):

Theorem 1. *The bivariate generating function of the probabilities is*

$$G(z,u) = \delta(u)\Psi\left(z\delta(u) + I(u)\right), \tag{19}$$

where Ψ is the function defined parametrically for $|u| < 1$ by

$$\Psi(I(u)) = J(u). \tag{20}$$

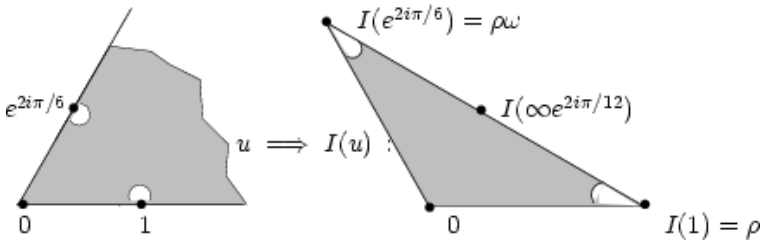


FIGURE 2. The "elementary triangle" T_0 (right) is the image of the basic sector S_0 (left) via the mapping $u \mapsto I(u)$.

2.3. **Dominant singularity.** The map $u \mapsto I(u)$ is analytic in the unit disk $|u| < 1$. Since $I(u)$ has nonnegative Taylor coefficients at 0, the image of the unit dis in the u-plane is a subset of the Dis centered at the origin and having radius ρ defined by $\rho = I(1)$. In particular, as $u \to 1^-$, one has $z \to \rho^-$. On the other hand the integral $J(u)$ diverges to $+\infty$ as $u \to 1^-$. There results that $z = \rho$ is a singularity of $\Psi(z)$. The quantity ρ happens to admit of closed form in terms of Gamma function factors (a complete Eulerian beta integral), and using Pringsheim's theorem, we get:

Lemma 1. *The function $\Psi(z)$ is analytic in the disco $|z| < \rho$, where*

$$\rho \equiv I(1) = \frac{1}{6} B\left(\frac{1}{6}, \frac{1}{3}\right) = \frac{1}{6}\frac{\Gamma(1/3)\Gamma(1/6)}{\Gamma(1/2)} \doteq 1.40218\,21053\,25454. \tag{21}$$

2.4. **The fundamental triangle.** An expansion of Ψ near 0 exhibits a periodicity of the coefficients of Ψ modulo 3. Local expansions of $I(u)$ and $J(u)$ show the limiting behaviors (in the sense of directional limits):

$$\Psi(z) \underset{z\to\rho}{\sim} \frac{1}{\rho - z}, \qquad \Psi(z) \underset{z\to\rho\omega}{\sim} \frac{1}{\rho\omega - z}, \qquad \Psi(z) \underset{z\to\rho\omega^2}{\sim} \frac{1}{\rho\omega^2 - z}.$$

Thus Ψ must be singular at the three points ρ, $\rho\omega$, and $\rho\omega^2$.

We consider $\delta(u) = (1 - u^6)^{1/6}$. Let $\zeta = e^{2i\pi/6}$ and define the region R_0 as the complex plane slit along the six rays $\zeta^j t$ with $t \in [1, +\infty[$, where $j = 0, \ldots, 5$. Let $R_0, \ldots, R_5$ be six copies of R_0 where by convention $\delta(t) \sim \zeta^j$ when $t \sim 0$ in R_j, with δ being also extended by continuity on R_j. By conveniently "gluing" together the six copies $R_0, \ldots, R_5$ along the rays, one obtain the Riemann Surface of δ onto which δ is single-valued.

We can restrict ourself by symmetry to the "slit" half-plane $\mathcal{H}$, where

$$\mathcal{H} := \{\, z : \Im(z) > 0 \vee (\Im(z) = 0 \wedge \Re(z) \geq 0) \,\}. \tag{22}$$

We then have:

Lemma 2. *The function $I(u)$ maps the interior of $(R_0 \cap \mathcal{H})$ conformally (i.e., in a one-to-one analytic manner) onto the interior of the equilateral triangle T with vertices $\rho, \rho\omega, \rho\omega^2$, where $\omega := e^{2i\pi/3}$.*

Figures 2 and 3 give hints of the proof that is omitted.

2.5. **Analytic continuation and elliptic connection.** The function Ψ is amenable to analytic continuation beyond its disk of convergence $|z| < \rho$. This can be done by rotating the fundamental triangle around $\rho, \rho\omega$ and $\rho\omega^2$. This is further extended to the whole complex plane punctured by the hexagonal lattice

$$\Lambda_{\text{hex}} := \left\{ n_1 e^{i\pi/6} + n_2 e^{i\pi/6} : n_1, n_2 \in \mathbb{Z} \right\}$$

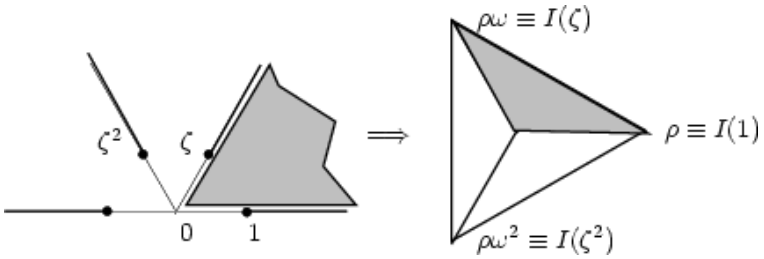


FIGURE 3. The "fundamental triangle" T (right) is the image of the slit upper half-plane ($R_0 \cap \mathcal{H}$) (left) via the mapping $u \mapsto I(u)$.

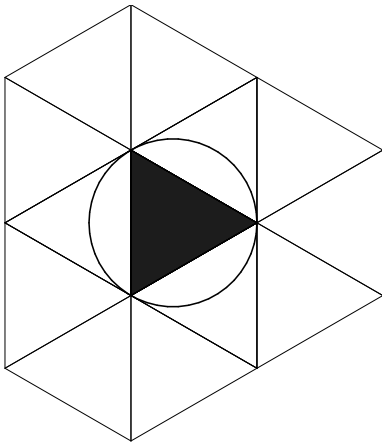


FIGURE 4. Rotated copies of the fundamental triangle around $\rho, \rho\omega, \rho\omega^2$ shown against the circle of convergence of $\Psi(z)$.

There we have:

Theorem 2. *The Ψ-function of the $\mathcal{T}_{2,3}$ model initialized with 2 balls of the first type is exactly*

$$\Psi(z) = \frac{1}{\rho\sqrt{3}}\left(-\zeta\left(\frac{z-\rho}{\rho\sqrt{3}}\right) + \zeta\left(-\frac{1}{\sqrt{3}}\right)\right), \qquad \textit{with} \quad \rho := \frac{1}{6}\,\frac{\Gamma(\frac{1}{3})\Gamma(\frac{1}{6})}{\Gamma(\frac{1}{2})}, \tag{23}$$

where $\zeta(z) := \zeta(z; \Lambda_{\text{hex}})$ is the Weierstraß zeta function of the hexagonal lattice defined by

$$\zeta(z; \Lambda_{\text{hex}}) := \frac{1}{z} + \sum_{w \in \Lambda_{\text{hex}} \setminus \{0\}} \left(\frac{1}{z-w} + \frac{1}{w} + \frac{z}{w^2}\right), \tag{24}$$

Sketch of proof: (1) each point z of the punctured points is reachable by a path $I(\gamma(u))$ for suitable γ; (2) poles and residues of both functions are the same; (3) Liouville theorem.

2.6. **Probabilistic consequences.** A consequence of the results of the preceding section is:

Corollary 1. *For the $\mathcal{T}_{2,3}$ model, the probability generating function $p_n(u) = \mathbf{E}(u^{X_n})$ admits an* exact *formula valid for all $n \geq 2$,*

$$p_n(u) = \sum_{n_1, n_2 = -\infty}^{+\infty} \left(K(u) + \frac{\rho\sqrt{3}}{\delta(u)}(n_1 e^{i\pi/6} + n_2 e^{-i\pi/6})\right)^{-n-1}, \tag{25}$$

where

$$K(u) := \frac{1}{\delta(u)} \int_u^1 \frac{t}{\delta(t)^5}\, dt, \qquad \delta(u) = (1-u^6)^{1/6}.$$

From this,(1) application of the Quasi-Powers Theorem of Bender and Hwang provides a Gaussian limit law, so as speed of convergence to the limit; (2) there are exact polynomial forms for the moments; (3) a large deviation law with exponential decay follows.

3. **General Case**

In the general case, the urn model with replacement is specified by the matrix

$$\begin{pmatrix} -a & a+s \\ b+s & -b \end{pmatrix},$$

with initial conditions: $a = a_0$ (black), $b = b_0$ (white).

The combinatorial analysis uses a "history" model with stamped urns; this leads to a PDE for the bivariate generating function $F(z,u)$ counting time and black balls, via a differential operator.

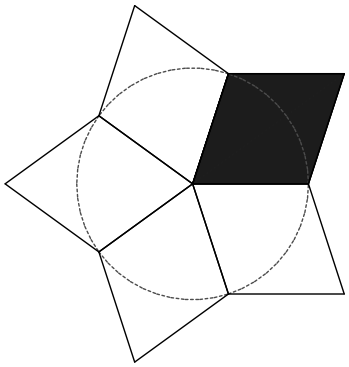


FIGURE 5. The fundamental polygon associated with the urn $(-1,4,4,-1)$.

The main features of the analysis are:

- For each system, there exists a regular fundamental polygon inside which the function Ψ is analytic.
- There are isolated singularities of Puiseux type on the boundary of the disk of convergence of $F(z,u)$ where u is considered as a parameter. This entails asymptotic equivalents for $[z^n]F(z,u) = p_n(u)$.
- The Quasi-Powers Theorem applies to $p_n(u)$, implying a Gaussian limit.
- The rth factorial moment of X_n is of hypergeometric type.
- The large deviation rate is fully characterized.

In general, the Puiseux singularities preclude solutions in terms of elliptic functions. This corresponds also to the impossibility of tiling the complex plan with the fundamental polygon (see for instance Figure 5 where $\Psi \asymp (\rho - z)^{-1/3}$).

The six elliptic urns are:

$$\begin{pmatrix} -2 & 3 \\ 4 & -3 \end{pmatrix}, \quad \begin{pmatrix} -1 & 2 \\ 3 & -2 \end{pmatrix}, \quad \begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix},$$
$$\begin{pmatrix} -1 & 3 \\ 3 & -1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 3 \\ 5 & -3 \end{pmatrix}, \quad \begin{pmatrix} -1 & 4 \\ 5 & -2 \end{pmatrix}.$$

Bibliography

[1] Flajolet (P.), Gabarró (J.), and Pekari (H.). – Analytic urns. – Submitted to *Annals of Probability*, available at `http://algo.inria.fr/flajolet/Publications/`.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 61–64.

Available online at the URL
http://algo.inria.fr/seminars/.

Analytic Urns of Triangular Form

Vincent Puyhaubert
Projet Algorithmes, INRIA (France)

September 22, 2003

Summary by Julien Fayolle

Abstract

The urn model was first introduced as a mathematical object by Eggenberger and Pólya [2] in 1923. It has since become handy to model some problems arising from computer science, like for instance the coupon collector problem, hashing problems or even random tree generation. The original goal of Pólya and Eggenberger was related to public health and the spread of epidemics.

Until recently the only approach to urn problems was through the probabilistic toolkit. In [3] Flajolet, Gabarró, and Pekari first devised an approach using analytic combinatorics for certain families of urn (balanced ones). Puyhaubert [4] travels down this same path and uses the characteristics method for solving partial differential equations to obtain more precise results on the asymptotic behavior of the urns (moments, limit law).

1. Model

An urn contains balls of different types, say colors. Suppose first that we have two types of balls, say black ones and white ones. The given of the problem is the initial configuration of the urn, i.e., the number of black and of white balls at the origin of time and a 2×2 *replacement* matrix $M = (m_{i,j})_{i,j}$ with integral entries that codes the evolution rules of the urn.

The time is discrete and at time n a ball is picked (and then put back in the urn) uniformly at random from the urn. If a black (resp. white) ball is drawn from the urn, we add $m_{1,1}$ (resp. $m_{2,1}$) black and $m_{1,2}$ (resp. $m_{2,2}$) white balls in the urn. For this talk the replacement matrix has fixed entries (for instance they do not depend on the time) and the sum of the entries of any row is constant (we say the urn is *balanced*). This balance condition means the number of balls in the urn at time $n+1$ does not depend on the ball picked at time n. Our matrices are triangular (say upper-triangular, meaning we necessarily have $m_{2,1} = 0$). One last condition is *tenability*: if the entries of the matrix were negative one might end up trying to remove balls that no longer exist in the urn, therefore the entries are all positive integers.

As an example, we explain the urn model for the coupon collector problem. We have n different coupons to collect. We code coupons already collected as black balls in the urn and those not already collected as white balls. At time zero, there are just n white balls in the urn for no coupon has yet been found. Each time we pick a coupon (or ball) either it is one we already possess in our collection (black ball) and then the urn composition does not change, or it is a new coupon (white ball) and then we remove the white ball and add a black ball to mark we have one more coupon in

our personal collection. We can model this behavior with the replacement matrix

$$M = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$$

and the initial composition of n white balls.

2. **Counting**

This part actually initiates from [3] and their results on analytic urns. Throughout the rest of the summary we use the replacement matrix M with an initial composition U_0

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}, \quad U_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

to serve as an example of the methods.

We enumerate all possible compositions of an urn (knowing its initial composition and its replacement matrix) at time n with a polynomial $f_n(u, v)$ where u marks the number of black balls and v the number of white balls in the urn. A composition can be counted several times if there are several histories (picks of the distinguished balls) leading to this composition at time n.

For our example we have $f_0(u, v) = uv$, $f_1(u, v) = u^2v^2+uv^3$, and $f_2(u, v) = 3uv^5+3u^2v^4+2u^3v^3$. Each history of length n (series of the n picks) has the same probability $2 \cdot 4 \cdot \ldots \cdot (2n)$.

From an urn composition $u^r v^s$ at time n there are r possibilities of picking a black ball and this leads to a composition $u^{r+1}v^{s+1}$ at time $n + 1$, and there are s possibilities of picking a white ball leading to a composition $u^r v^{s+2}$ at time $n+1$. Hence from the $u^r v^s$ composition at time n we have $ru^{r+1}v^{s+1} + su^r v^{s+2}$ possible compositions at time $n + 1$. We introduce a differential operator Γ to code the evolution of the monomials $u^r v^s$:

$$\Gamma = uv \cdot u\frac{\partial}{\partial u} + u^0v^2 \cdot v\frac{\partial}{\partial v}. \tag{1}$$

This operator is linear so it extends to the whole polynomial allowing us to describe in a simple manner the possible evolutions of the urn composition: $f_{n+1}(u, v) = \Gamma f_n(u, v)$.

We introduce the main generating function for our study:

$$H(z, u, v) = \sum_{n\geq 0} f_n(u, v)\frac{z^n}{n!},$$

it satisfies the initial condition $H(0, u, v) = uv$ and the partial differential equation $\frac{\partial}{\partial z}H = \Gamma H$. We know there is one sole generating function satisfying these two conditions, and from it we can derive the moments of the random variable X_n counting the number of black balls in the urn and also its limit law.

Why focus uniquely on black balls? Because the balance condition on the urn means the population growth of the urn is deterministic: whatever the history of the urn the population increases by two at every replacement. It leads to $s + r = 2 + 2n = 2(n + 1)$ balls in the urn at time n.

3. **A Method for Solving PDEs**

The partial differential equation

$$u^2v\frac{\partial H}{\partial u} + v^3\frac{\partial H}{\partial v} = \frac{\partial H}{\partial z} \tag{2}$$

can be solved using the characteristics method (see [1] for a thorough course on it). We first introduce the notion of first integral and some of its properties before applying them to our simple example.

Definition 1. Let $x'(t) = f(x)$ be an ordinary differential equation where x is a $\mathcal{C}^1$ function from an open I of $\mathbb{R}$ to an open U of $\mathbb{R}^n$ and f is a $\mathcal{C}^1$ function from U to $\mathbb{R}^n$. A $\mathcal{C}^1$ function ψ from U to $\mathbb{R}$ is said to be a first integral for the equation $x'(t) = f(x)$ if for any solution $x(t)$ to the differential equation $\psi(x(t))$ is constant.

The equation $x'(t) = f(x)$ is actually a system of n ODEs if we look at the coordinates $(x_i)_i$ and $(f_i)_i$ of the functions $x(t)$ and $f(x)$ on the canonical basis of $\mathbb{R}^n$. The benefit of using prime integrals is to avoid partial derivatives to deal solely with ODEs. The next proposition describes this link between ODEs and PDEs, its proof is quite straightforward.

Proposition 1. *A function ψ is a prime integral of $\frac{dx}{dt}(t) = f(x)$ if and only if for any point of U we have*

$$\sum_{i=1}^{n} f_i(x_1, \ldots, x_n) \frac{\partial \psi}{\partial x_i}(x_1, \ldots, x_n) = 0. \tag{3}$$

For a "good" differential system of n equation there exist $n-1$ prime integrals $(\psi_i)_i$ whose derivatives are linearly independent such that any prime integral can be expressed as $\phi(\psi_1, \ldots, \psi_{n-1})$ for any arbitrary $\mathcal{C}^1$ function ϕ.

For our guideline example we have to solve the PDE (2) which is of the type of (3). Given the proposition we only have to look for prime integrals of the system of three equations

$$\frac{du}{dt} = u^2 v, \qquad \frac{dv}{dt} = v^3, \quad \text{and} \quad \frac{dz}{dt} = -1. \tag{4}$$

We eliminate the dt from all three equations and obtain a system of two equations

$$\frac{dv}{v^3} = -dz, \quad \text{and} \quad \frac{du}{u^2} = \frac{dv}{v^2},$$

then we integrate and get two first integrals (checking it is easy) whose derivatives are linearly independent:

$$\psi_1(z, u, v) = z - \frac{1}{2v^2}, \quad \text{and} \quad \psi_2(z, u, v) = \frac{1}{v} - \frac{1}{u}.$$

The generating function for the urn is written in the form

$$H(z, u, v) = \phi\left(z - \frac{1}{2v^2}, \frac{1}{v} - \frac{1}{u}\right)$$

for some arbitrary function ϕ. We also have an initial condition $H(0, u, v) = uv$ that leads to

$$H(z, u, v) = uv(1 - 2v^2 z)^{-1/2} \left(1 - uv^{-1}(1 - (1 - 2v^2 z)^{1/2})\right)^{-1}.$$

4. **Results**

We now take a more general 2×2 triangular replacement matrix M and an initial composition (a_0, b_0) for the urn, thus with an initial population of $t_0 = a_0 + b_0$ balls.

$$M = \begin{pmatrix} a & b-a \\ 0 & b \end{pmatrix}$$

We solve the PDE

$$u^{a+1} v^{b-a} \frac{\partial H}{\partial u} + v^{b+1} \frac{\partial H}{\partial v} = \frac{\partial H}{\partial z}$$

associated to the urn using the characteristics method from the previous section. It leads to the generating function

$$H(z,u,v) = u^{a_0} v^{b_0} (1 - bv^b z)^{-b_0/b} (1 - u^a v^{-a} (1 - (1 - bv^b z)^{a/b}))^{-a_0/a}. \tag{5}$$

If we denote by X_n the random variable counting the number of black balls at time n, we obtain the probability of having $a_0 + ka$ black balls at time n and also a full expansion of its lth-order moment.

$$\mathbf{P}(X_n = a_0 + ka) = \frac{n!}{t_0 \cdots (t_0 + (n-1)s)} \binom{\frac{a_0}{a} + k - 1}{k} \sum_{i=0}^{k} (-1)^i \binom{k}{i} \binom{\frac{b_0 - ai}{b} + n - 1}{n} b^n \tag{6}$$

$$\mathbf{E}((X_n)^l) = a^l \frac{\Gamma(\frac{a_0+la}{a})\Gamma(\frac{t_0}{b})}{\Gamma(\frac{a_0}{a})\Gamma(\frac{t_0+la}{b})} n^{la/b} + O(n^{(l-1)a/b}). \tag{7}$$

We are also interested in the local limit law of the random variable X_n. The density can be expressed using Mittag–Leffler functions so that the limit law is a stable law lookalike. More precisely, for any positive x such that $xn^{a/b}$ is an integer we have the equivalent

$$\mathbf{P}(X_n = a_0 + axn^{a/b}) = \frac{1}{n^{a/b}} \frac{\Gamma(\frac{t_0}{b})}{\Gamma(\frac{a_0}{a})} x^{\frac{a_0}{a}-1} \sum_{k \geq 0} \frac{(-1)^k}{\Gamma(\frac{b_0 - ka}{b})} \frac{x^k}{k!} + O\left(\frac{1}{n^{2a/b}}\right).$$

5. **Triangular Urns of Size 3**

We can easily adapt the generating function approach to urns filled with balls of three distinct colors. The matrix is then 3×3. We still ask for a triangular and balanced matrices. The methods are the same as for the two color case but they are original contributions from Puyhaubert since the probabilistic method needs additional constraints (irreducibility) to deal with 3×3 matrices.

Balanced 3×3 triangular replacement matrices (for balance 2 and 3) are classified according to the asymptotic growth of their three populations.

The generating function methodology leads to results for matrices of arbitrary size but the computations, especially to determine first integrals become heavier.

Bibliography

[1] Cartan (Henri). – *Calcul différentiel*. – Hermann, 1967.

[2] Eggenberger (Florian) and Pólya (George). – Über die Statistik verketteter Vorgänge. *Zeitschrift für angewandte Mathematik und Mechanik*, vol. 1, 1923, pp. 279–289.

[3] Flajolet (Philippe), Gabarró (Joaquim), and Pekari (Helmut). – Analytic urns. *Annals of Probability*, 2005.

[4] Puyhaubert (Vincent). – *Modèles d'urnes et phénomènes de seuil en combinatoire analytique*. – PhD thesis, École polytechnique, 2005.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 65–70.

Available online at the URL
http://algo.inria.fr/seminars/.

Suffix Trees and Simple Sources

Julien Fayolle
Algorithms Project, Inria (France)

November 17, 2003

Summary by Pierre Nicodème

Abstract

Using an intricate method, Jacquet and Szpankowski [2] compared the depth of insertion into suffix-trees and tries in the non-uniform Bernoulli model, as well as the average size of suffix-trees and tries under the same model. They proved that the depth of insertion has asymptotically the same probabilistic behaviour in both cases, and that the average sizes of a trie and a suffix-tree built with n keys are asymptotically equivalent. Julien Fayolle uses a simpler combinatorial approach to compare both tree structures. When considering a two-letters alphabet with letters probability p and $q = 1 - p$, he improves the asymptotic estimation for the expectations of external path length and size of the suffix-tree (more specifically, he obtains an asymptotic bound $O(n^{0.85})$ for the difference of the two expectations when $p \in [0.46, 0.54]$). The Lempel–Ziv compression algorithm and its variants use suffix-trees as underlying data-structure.

1. Introduction

We consider a memoryless source over an alphabet $\Sigma = \{\mathtt{0}, \mathtt{1}\}$, with $\mathbf{P}(\mathtt{0}) = p, \mathbf{P}(\mathtt{1}) = 1 - p = q$ and $p > q$.

Trie. Let X be a finite set of infinite words over Σ. A trie with input keys X is defined by

$$\mathrm{trie}(X) = \begin{cases} \emptyset & \text{if } |X| = 0, \\ \bullet & \text{if } |X| = 1, \\ \langle \bullet, \mathrm{trie}(X\backslash\mathtt{0}), \mathrm{trie}(X\backslash\mathtt{1}) \rangle & \text{elsewhere,} \end{cases}$$

where the symbol $\bullet$ represents a node and $X\backslash a = \{\, u : (w \in X \text{ and } w = au) \,\}$, for $a \in \Sigma$.

Suffix tree. The suffix-tree of n keys over an infinite random string Y is the trie built over the set X of the n first suffixes of Y.

Definitions. For any string $\omega \in \Sigma^\star$, let N_ω be the *number of keys of* X (or first n suffixes of Y) whose prefix is ω.

Given a trie $\mathcal{T}$, we only consider internal nodes; therefore,

- for any internal node ν of $\mathcal{T}$, if ω_ν is the word spelled by reading the labels of the edges of $\mathcal{T}$ from the root to ν, we have $N_{\omega_\nu} \geq 2$. We write $\omega_\nu \in \mathcal{T}$ in this case;
- reciprocally, if $N_\omega = 0$, there is no node in $\mathcal{T}$ accessed by reading ω and if $N_\omega = 1$, ω leads to a leaf; ($\omega \notin \mathcal{T}$ in both cases).

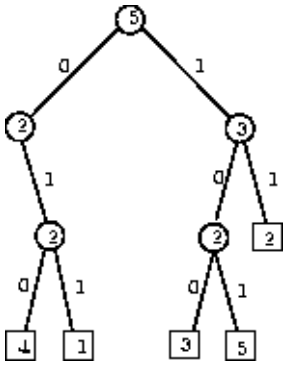


FIGURE 1. Trie for the keys `011...`, `110...`, `101...`, `010...` and `100...`; this is also the suffix tree built with the five first suffixes of the sequence `0110100...`. Each internal node is numbered with the cardinality of leaves in its subtree; each leaf is numbered as starting position of the corresponding suffix in `0110100...`.

Probabilistic model. In what follows, any random variable is understood as conditionned by a trie $\mathcal{T}$ over n keys, this tree being built either over a random set of infinite keys X, or over a random infinite sequence Y (suffix-tree). The generating function $\mathcal{L}(z)$ of a language $\mathcal{L}$ will always be considered in the weighted case, as $\mathcal{L}(z) = \sum_{\omega\in\mathcal{L}} \mathbf{P}(\omega) z^{|\omega|}$. We note in the following $\pi_\omega = \mathbf{P}(\omega)$ for all words ω.

External path length. Let L denote the external path length of a trie $\mathcal{T}$ build over a set of keys X. As clearly seen on the example of Figure 1, we have

$$L = \sum_{\omega\in\mathcal{T}} N_\omega \mathbf{1}_{\{N_\omega\geq 2\}} = \sum_{\omega\in\Sigma^\star} N_\omega \mathbf{1}_{\{N_\omega\geq 2\}},$$

which follows from the fact that $\mathbf{1}_{\{N_\omega\geq 2\}} = 0$ for $\omega \notin \mathcal{T}$.

Remarking that $\mathbf{E}(N_\omega \mathbf{1}_{\{N_\omega=0\}}) = 0$, this gives

(1)

$$\mathbf{E}(L) = \mathbf{E}\left(\sum_{\omega\in\Sigma^\star} N_\omega \mathbf{1}_{\{N_\omega\geq 2\}}\right) = \sum_{\omega\in\Sigma^\star}\Big(\mathbf{E}(N_\omega) - \mathbf{E}\left(N_\omega \mathbf{1}_{\{N_\omega=1\}}\right)\Big) = \sum_{\omega\in\Sigma^\star}\Big(\mathbf{E}(N_\omega) - \mathbf{P}(N_\omega = 1)\Big).$$

When $|X| = n$, we note respectively $\mathbf{E}^{(S_n)}(L)$ and $\mathbf{E}^{(T_n)}(L)$ the expectations of the external path length of a suffix-tree and of a trie. We will consider in the following $\mathbf{E}_L^{(S)}(z) = \sum_n \mathbf{E}^{(S_n)}(L) z^n$. We write in the same spirit $\mathbf{P}^{(S_n)}(N_\omega = 1), \mathbf{P}^{(T_n)}(N_\omega = 1)$, and define $\mathbf{E}_{N_\omega}^{(S)}(z) = \sum_n \mathbf{E}^{(S_n)}(N_\omega) z^n$ and $\mathbf{P}_{(N_\omega=1)}^{(S)}(z) = \sum_n \mathbf{P}^{(S_n)}(N_\omega = 1) z^n$.

Size. We obtain similarly for the size S of a trie $\mathcal{T}$

$$S = \sum_{\omega\in\mathcal{T}} \mathbf{1}_{\{N_\omega\geq 2\}} = \sum_{\omega\in\Sigma^\star} \mathbf{1}_{\{N_\omega\geq 2\}} \Rightarrow \mathbf{E}(S) = \sum_{\omega\in\Sigma^\star} \mathbf{P}(N_\omega \geq 2) = \sum_{\omega\in\Sigma^\star} \big(1 - \mathbf{P}(N_\omega = 0) - \mathbf{P}(N_\omega = 1)\big).$$

2. Trie: External Path Length

The external path length and size of a trie over n keys have been extensively studied. We have for a trie over n keys (see [1] for a proof; proofs for size and external path length are similar).

Theorem 1. *Asymptotically, for a memoryless source (p, q), the expectation of the external path length L is:*

- *if the source is periodic ($\log p/\log q$ rational)*

$$\mathbf{E}^{(T_n)}(L) = -\frac{n\log n}{p\log p + q\log q} + Kn + n\epsilon(n) + o(n),$$

  where $\epsilon(n)$ is a periodic function of weak amplitude;
- *elsewhere (aperiodic source)*

$$\mathbf{E}^{(T_n)}(L) = -\frac{n\log n}{p\log p + q\log q} + Kn + o(n)$$

3. Suffix Tree: External Path Length

For any ω, the probability of occurrence of ω as one of the first n suffixes of the string Y is independent of the position; therefore $\mathbf{E}^{(S_n)}(N_\omega) = n \times \pi_\omega$.

We similarly have $\mathbf{E}^{(T_n)}(N_\omega) = n \times \pi_\omega = \mathbf{E}^{(S_n)}(N_\omega)$.

We want to compute the probability that ω occurs once and only once in the string. We use a variation of Guibas and Odlyzko's method. (See [3] p. 374).

We consider the autocorrelation set $\mathcal{A}_\omega$ of ω, defined as

$$\mathcal{A}_\omega = \left\{\, h : \omega.h = u.\omega \text{ and } |h| < |\omega| \,\right\}.$$

Let (a) $\mathcal{F}_\omega$, (b) $\mathcal{T}_\omega$ and (c) $\mathcal{W}_\omega$ be respectively the language of texts (a) whose first and lone occurrence of ω is at the end of the text (*First* occurrence in a text), (b) whose concatenation to ω do not create a new occurrence of ω (*Tail* following the last occurrence of ω) and (c) *Without* occurrence of ω. Let σ be any letter of the alphabet Σ.

We have two formal equations

$$\mathcal{W}_\omega.\sigma = \mathcal{W}_\omega + \mathcal{F}_\omega - \epsilon \quad \text{and} \quad \mathcal{W}_\omega.\omega = \mathcal{F}_\omega.\mathcal{A}_\omega,$$

where ϵ is the empty word (that also belongs to $\mathcal{A}_\omega$). Products and unions are unambiguous, and we obtain for the weighted generating functions (see Section 1)

$$\mathcal{W}_\omega(z) \times z = \mathcal{W}_\omega(z) + \mathcal{F}_\omega(z) - 1 \quad \text{and} \quad \mathcal{W}_\omega(z) \times \pi_\omega z^{|\omega|} = \mathcal{F}_\omega(z) \times \mathcal{A}_\omega(z).$$

Solving for $\mathcal{W}_\omega(z)$ and $\mathcal{F}_\omega(z)$, we obtain

$$\mathcal{F}_\omega(z) = \frac{\pi_\omega z^{|\omega|}}{\pi_\omega z^{|\omega|} + (1-z)\mathcal{A}_\omega(z)}.$$

Let $\leftarrow$ denote backwards reading of words of a language. For any $\mathcal{L}$, by reading backwards and forwards words, we have a bijection between $\mathcal{L}$ and $\overleftarrow{\mathcal{L}}$. This implies (with a memoryless source) that $\mathcal{L}(z) = \overleftarrow{\mathcal{L}}(z)$. But we have $\mathcal{W}_\omega(z) = \overleftarrow{\mathcal{W}_\omega}(z) = \overleftarrow{\mathcal{W}_{\overleftarrow{\omega}}}(z) = \mathcal{W}_{\overleftarrow{\omega}}(z)$ and therefore $\mathcal{F}_\omega(z) = \mathcal{F}_{\overleftarrow{\omega}}(z)$. This implies that

$$\mathcal{F}_\omega = \overleftarrow{\overleftarrow{\omega}.\mathcal{T}_{\overleftarrow{\omega}}} \qquad \Rightarrow \qquad \mathcal{T}_{\overleftarrow{\omega}}(z) = \frac{\overleftarrow{\mathcal{F}_\omega}(z)}{\overleftarrow{\omega}(z)} = \frac{\overleftarrow{\mathcal{F}_{\overleftarrow{\omega}}}(z)}{\omega(z)} = \mathcal{T}_\omega(z) = \frac{\mathcal{F}_\omega(z)}{\omega(z)} = \frac{\mathcal{F}_\omega(z)}{\pi_\omega z^{|\omega|}}.$$

We also have $\mathcal{O}_\omega(z) = \mathcal{F}_\omega(z)\mathcal{T}_\omega(z)$ for the generating function $\mathcal{O}_\omega(z)$ of texts with exactly one match with ω. Summing up over ω, we obtain the generating function $E_L^{(S)}(z)$ of expectations of

the external path length of a suffix-tree (see Jacquet and Szpankowski [2] for another proof)

$$\mathbf{E}_L^{(S)}(z) - \sum_{\omega\in\Sigma^\star} \mathbf{E}_{N_\omega}^{(S)}(z) = \sum_{\omega\in\Sigma^\star} \mathbf{P}_{(N_\omega=1)}^{(S)}(z) = \sum_{\omega\in\Sigma^\star} \mathcal{O}_\omega(z) = \sum_{\omega\in\Sigma^\star} \frac{\pi_\omega z^{|\omega|}}{\left(\pi_\omega z^{|\omega|} + (1-z)\mathcal{A}_\omega(z)\right)^2}. \tag{2}$$

4. External Path Length, Suffix Tree versus Trie

We consider (the index n corresponding to n keys) the differences

$$\Delta_n = \mathbf{E}^{(T_n)}(L) - \mathbf{E}^{(S_n)}(L) = \sum_{\omega\in\Sigma^\star} \delta_\omega^{(n)} = \sum_{\omega\in\Sigma^\star} \mathbf{P}^{(T_n)}(N_\omega = 1) - \mathbf{P}^{(S_n)}(N_\omega = 1), \tag{3}$$

(since $\mathbf{E}^{(T_n)}(N_\omega) = \mathbf{E}^{(S_n)}(N_\omega)$ for all ω). We will prove that $\Delta_n = O(n^{0.85})$ for $0.5 < p < 0.54$.

The minimum μ_n of the fillup levels of both trees is $\alpha\log(n)$ for a given $\alpha > 0$ with probability one; all the following will be conditionned by the fact that $\mu_n = \alpha\log(n)$. With $|\omega| < \mu_n$, we have $\delta_\omega^{(n)} = 0$ and when $|\omega| \geq \mu_n$, asymptotically, $\pi_\omega = o(1)$.

4.1. Asymptotic contribution of the trie.

As a consequence of the preceding remark, we have

$$\mathbf{P}^{(T_n)}(N_\omega = 1) = n\pi_\omega \times (1-\pi_\omega)^{n-1} \simeq n\pi_\omega \times e^{-n\pi_\omega}. \tag{4}$$

4.2. Asymptotic contribution of the suffix-tree.

Each function $\mathcal{O}_\omega(z)$ in Equation 2 is a rational function with dominant pole ρ_ω. We begin by isolating the dominant poles of these functions that give the asymptotic behaviour of the terms corresponding to the suffix-tree in Δ_n.

Lemma 1. *For $1 < R < 1/p$ and $|\omega| \geq \mu_n$, the set of poles of the functions $\mathcal{O}_\omega(z)$ contained in the disk centered at the origin and of radius R is exactly $\{\rho_\omega;\ \omega \in \Sigma^\star, |\omega| \geq \mu_n\}$.*

Proof. The proof is based on an application of the Rouché theorem; let $f(z) = \pi_\omega z^{|\omega|}$, and $g(z) = (1-z)\mathcal{A}_\omega(z)$. We are above the level μ_n and therefore, for $|z| < 1/p$, we have $|f(z)| = o(1)$. Let d be the smallest period of ω.

– If $d < |\omega|/2$, we have $\omega = u^r.v$, (with $|v| < |u| = d$ and v a prefix of u), and the second smallest period is at least $|\omega|/2$ (a consequence of the Wilf theorem). This gives (omitting the subscript ω)

$$\mathcal{A}(z) = 1 + S(z) + T(z), \qquad \text{with} \qquad S(z) = \pi_u z^{|u|} + \cdots + (\pi_u z^{|u|})^r,$$

where $T(z)$ is a polynomial of lowest degree $\geq \mu_n/2$; this implies $|T(z)| = o(1)$ for $|z| < 1/p$ and

$$|\mathcal{A}(z)| = \left|\frac{1}{1-\pi_u z^{|u|}}\right| + o(1).$$

For all $d \geq 2$, this implies (up to negligible terms)

$$|\mathcal{A}(z)| \geq \frac{1}{1+\pi_u|z|^{|u|}} \geq \frac{1}{1+p|z|} \qquad \text{for} \qquad |z| < \frac{1}{p}.$$

We have the same lower bound for $d = 1$.

– If $d > |\omega|/2$, we have $S(z) = 0$ and $|\mathcal{A}(z)| = 1 + o(1)$ for $|z| < 1/p$.

Therefore, for any $R < 1/p$ (and we choose in the following $R > 1$), there exists a number N such that for $n > N$, on the circle $|z| = R$, we have $|\mathcal{A}_\omega(z)| > \kappa$ for a given $\kappa > 0$ and for all ω such that $|\omega| > \mu_n$; this implies $|f| < |g|$ over this circle. Moreover f and g are analytic everywhere, which implies that $f + g$ has as many zeros as g inside the disk $\mathcal{D}_R = \{z, |z| < R\}$, for any ω. The polynomial $\mathcal{A}(z)$ has no roots inside the disk $|z| < R$ and therefore $f(z) + g(z)$ has only one root inside the disk $\mathcal{D}_R$. □

For each ω, we compute

$$o_n^{(\omega)} = [z^n]\mathcal{O}_\omega(z) = \operatorname{Res}\left(\frac{\mathcal{O}_\omega(z)}{z^{n+1}}, 0\right) = I(\mathcal{C}_R) - \operatorname{Res}\left(\frac{\mathcal{O}_\omega(z)}{z^{n+1}}, \rho_\omega\right), \quad \text{where} \quad I(\mathcal{C}_R) = \int_{\mathcal{C}_R} \frac{\mathcal{O}_\omega(z)}{z^{n+1}} dz,$$

and $\mathcal{C}_R$ is the circle $|z| = R$. Considering

$$\mathcal{O}_\omega(z) = \frac{\pi_\omega z^{|\omega|}}{\left(\pi_\omega z^{|\omega|} + (1-z)\mathcal{A}_\omega(z)\right)^2},$$

for $|\omega| > \mu_n$ and $|z| = R$ ($R \in]1, 1/p[$) we have $\pi_\omega z^{|\omega|} = o(1)$ and $|(1-z)\mathcal{A}(z)| \geq (R-1)\kappa$. Therefore $I(\mathcal{C}_R) = O(R^{-n})$.

By bootstrapping, we have $\rho_\omega = 1 + \pi_\omega / A_\omega(1) + o(\pi_\omega)$. An expansion of the denominator of $\mathcal{O}_\omega$ in a neighborhood of ρ_ω gives

$$\mathbf{P}^{(S_n)}(N_\omega = 1) = o_n^{(\omega)} = n\pi_\omega e^{-\frac{n\pi_\omega}{A_\omega(1)}} + O\left(n\pi_\omega^2 e^{-\frac{n\pi_\omega}{A_\omega(1)}}\right) + O(|\omega| n\pi_\omega^2) + C\left(\frac{1}{R}\right)^n. \tag{5}$$

Plugging Equations 4 and 5 into Equation 3 gives

$$\Delta_n = \sum_{\omega \in \Sigma^\star, |\omega| > \mu_n} \delta_\omega^{(n)} \quad \text{with} \quad \delta_\omega^{(n)} \simeq \sum_{\omega \in \Sigma^\star, |\omega| > \mu_n} n\pi_\omega \left(e^{-n\pi_\omega / A_\omega(1)} - e^{-n\pi_\omega}\right). \tag{6}$$

4.3. **Bounding Δ_n by partitioning the motifs ω.** The aim of this section is to prove the validity of each entry of the following table (where $C_p = \log 2 / \log(1/p)$ and $\Omega_s^{(n)}, \Omega_{i,p}^{(n)}, \Omega_{i,a}^{(n)}$, and $\Omega_l^{(n)}$ respectively are the set of short, intermediate periodic, intermediate aperiodic and long motifs when the number of input keys is n).

short patterns $\lvert\omega\rvert < \frac{5}{6}\log_{1/q} n$	intermediate patterns	long patterns $1.5\log_{1/p} n < \lvert\omega\rvert$
$\Delta_n^{(s)} = \sum_{\omega \in \Omega_s^{(n)}} \delta_\omega^{(n)} = o(1)$	periodic ($A_\omega(1) \geq 1 + 2^{-\lvert\omega\rvert/2}$) $\Delta_n^{(i,p)} = \sum_{\omega \in \Omega_{i,p}^{(n)}} \delta_\omega^{(n)} = O\left(n^{0.75 C_p} \log n\right)$	$\Delta_n^{(l)} = \sum_{\omega \in \Omega_l^{(n)}} \delta_\omega^{(n)} = O\left(\sqrt{n}\right)$
($\mu_n < \lvert\omega\rvert$)	aperiodic ($A_\omega(1) < 1 + 2^{-\lvert\omega\rvert/2}$) $\Delta_n^{(i,a)} = \sum_{\omega \in \Omega_{i,a}^{(n)}} \delta_\omega^{(n)} = O\left(n^{0.75 C_p}\right)$	

4.3.1. *Short patterns.* For these patterns, we have

$$n\pi_\omega > n q^{\frac{5}{6}\log_{1/q} n} = n^{1-\frac{5}{6}} = n^{1/6} \to \infty \qquad \text{and} \qquad \left|\Omega_s^{(n)}\right| = O(n^\alpha \log n) \quad \text{where } \alpha = \frac{5\log 2}{6\log(1/q)}.$$

Therefore $\Delta_n^{(s)}$ behaves as $n^\alpha \log n \times e^{-n^{1/6}}$ as n tends to infinity and is $o(1)$.

4.3.2. *Long patterns.* In this case, let $k_l(n) = 1.5\log_{1/p} n$; we have

$$n\pi_\omega \leq n p^{k_l(n)} = n p^{1.5\log_{1/p} n} = n^{-0.5} \to 0.$$

Expanding $\delta_\omega^{(n)}$ for small np_ω gives $\delta_\omega^{(n)} \simeq (n\pi_\omega)^2 (1 - 1/A_\omega(1))$. Therefore we have

$$\Delta_n^{(l)} \simeq \sum_{k \geq k_l(n)} \sum_{\omega \in \Sigma^k} n^2 \pi_\omega^2 \left(1 - \frac{1}{A_\omega(1)}\right) \quad \text{and} \quad \sum_{\omega \in \Sigma^k} \pi_\omega^2 = \sum_{0 \leq i \leq k} \binom{k}{i} p^{2i} q^{2(k-i)} = (p^2 + q^2)^k < p^k.$$

This implies $\Delta_n^{(l)} = O(n^2 p^{k_l(n)}) = O(\sqrt{n})$.

4.3.3. *Intermediate patterns.* Julien Fayolle proves in [1] that $\sum_{\omega\in\Sigma^k} A_\omega(1) = 2^k + k - 1$. He defines the set of intermediate *periodic* patterns as

$$\Omega_{i,p}^{(n)} = \left\{\omega :\ k_s(n) < |\omega| \le k_l(n),\ A_\omega(1) \ge 1 + 2^{-k/2}\right\},$$

where $k_s(n) = 5/6 \log_{1/q} n$ (in the uniform case these patterns verify $d < |\omega|/2$). He also proves that $|\Omega_{i,p}^{(n)}| \le k 2^{k/2}$.

Summing up for the intermediate periodic patterns, we obtain

$$\Delta_n^{i,p} = \sum_{\omega\in\Omega_{i,p}^{(n)}} \delta_\omega^{(n)} < K \sum_{k=k_s(n)}^{k_l(n)} k 2^{k/2} = O\left(\log n \times e^{\frac{1.5 \log 2 \log n}{2\log(1/p)}}\right) = O(n^{0.85}) \quad \text{for} \quad p < 0.54.$$

The set $\Omega_{i,a}^{(n)}$ of intermediate *aperiodic* patterns is the complementary set of $\Omega_{i,p}^{(n)}$, inside the bounds $k_s(n) < |\omega| < k_l(n)$. We therefore have $1/A_\omega(1) \ge 1/(1 + 2^{-|\omega|/2}) \ge 1 - 2^{-|\omega|/2}$ and for $|\omega| = k$

$$\delta_\omega^{(n)} \le n\pi_\omega \left(e^{n\pi_\omega 2^{-k/2}} - 1\right).$$

We also have for $\omega \in \Omega_{i,a}^{(n)}$

$$n\pi_\omega 2^{-|\omega|/2} < np^{k_s(n)} 2^{-k_s(n)/2} \to 0 \quad \text{for} \quad \frac{5}{6}\frac{\log(p/\sqrt{2})}{\log(1/q)} + 1 < 0 \quad \text{or} \quad p < p_0 = 0.5469.$$

By expanding the exponential in the neighborhood of zero, we get $\delta_\omega^{(n)} \simeq (n\pi_\omega)^2 e^{-n\pi_\omega} 2^{-|\omega|/2}$, which gives (remarking that $x^2 e^{-x}$ is bounded on $\mathbb{R}^+$)

$$\Delta_n^{(i,a)} \le \sum_{k=k_s(n)}^{k_l(n)} K' 2^k 2^{-k/2} = O\left(e^{\frac{1.5 \log 2 \log n}{2\log(1/p)}}\right) = O\left(n^{0.85}\right) \quad \text{for} \quad p < p_0.$$

Remark that we also have $\Delta_n^{(i,p)} = O\left(n^{0.85}\right)$ for $p < p_0$.

4.4. **End result.** Summarizing the preceding results gives

Theorem 2. *For a suffix-tree with n keys, we have asymptotically for $p \le 0.54$*

$$\mathbf{E}(L_n^{(S)}) = \frac{n \log n}{p \log p + q \log q} + (K + \epsilon(n))n + O\left(n^{0.85}\right),$$

where $L_n^{(S)}$ is the external path length of the tree and $\epsilon(n)$ is a periodic function of small amplitude.

The same method applies for analysis of the asymptotic expectation of size.

Bibliography

[1] Fayolle (Julien). – Paramètres des arbres suffixes dans le cas de sources simples. – 2003. Master's thesis.

[2] Jacquet (P.) and Szpankowski (W.). – Autocorrelation on words and its applications. Analysis of Suffix Trees by String Ruler Approach. *Journal of Combinatorial Theory, Series A*, n° 66, 1994, pp. 237–269.

[3] Sedgewick (Robert) and Flajolet (Philippe). – *An Introduction to the Analysis of Algorithms.* – Addison-Wesley Publishing Company, 1996, 512p.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 71–74.

Available online at the URL
http://algo.inria.fr/seminars/.

Efficient Computation of a Class of Continued Fraction Constants

Loïck Lhote
GREYC, Université de Caen (France)

November 17, 2003

Summary by Julien Clément

Abstract

There are numerous instances where mathematical constants do not admit a closed form. It is then of great interest to compute them, possibly in an efficient way. So the question is: does there exist an algorithm that computes the first d-digits of the constants and if so, what is the complexity in the number of arithmetic operations? We recall that a constant is said to be *polynomial-time computable* if its first d digits can be obtained with $O(d^r)$ arithmetic operations. Here we consider a particular class of constants arising in the field of the dynamical analysis of algorithms and dynamical systems. The constants to compute are of a "spectral" nature since they are closely related to the spectrum of some transfer operators.

1. Dynamical Systems and Constants

Dynamical analysis of algorithms was introduced by Brigitte Vallée [10] and gives a general framework to study complex and realistic systems. Roughly speaking, the idea is to track the execution of an algorithm through trajectories in its associated dynamical system. The techniques are based upon functional analysis and generating operators. This approach has been successfully applied to euclidean algorithms and yields very precise (and new) results. Viewing the dynamical system as a way to produce symbols, this framework allows also to study data structures built upon words like digital trees and tries [1, 9].

In the general and basic setting [6, 7], we are interested in complete dynamical systems $(\mathcal{I}, T)$ formed with an interval $\mathcal{I}$ and a map $T : \mathcal{I} \to \mathcal{I}$ which is piecewise surjective and of class $\mathcal{C}^2$. We denote by $\mathcal{G}$ the set of the inverse branches of T; then, $\mathcal{G}^k$ is the set of the inverse branches of T^k. It is known that contraction properties of the inverse branches are essential to obtain "good" properties on the dynamical system. Usually, what is needed is the existence of a disk D which is strictly mapped inside itself by all the inverse branches $h \in \mathcal{G}$ of the system (i.e., $h(\overline{D}) \subset \overline{D}$).

If f_0 is an initial density on $\mathcal{I}$, repeated applications of the map T modify the density and the successive densities $f_1, f_2, \ldots, f_n, \ldots$ describes the global evolution of the system at time $t = 0, 1, 2 \ldots, n, \ldots$. The operator $\mathbf{G}$ such that $f_1 = \mathbf{G}[f_0]$ and more generally $f_n = \mathbf{G}[f_{n-1}] = \mathbf{G}^n[f_0]$ for all n is called the density transformer. It is defined as

$$\mathbf{G}[f](z) = \sum_{h \in \mathcal{G}} |h'(z)| \; f \circ h(z)$$

It acts on the functional space $A_\infty(D)$ for some convenient disk D where $\mathcal{I} \subset D$ and

$$A_\infty(D) := \{f : D \to \mathbb{C}; f \text{ analytic on } D \text{ and continuous on } \overline{D}\}.$$

A perturbation of the density transformer, the transfer operator $\mathbf{G}_s$, defined as

$$\mathbf{G}_s[f](z) = \sum_{h\in\mathcal{G}} |h'(z)|^s \; f\circ h(z)$$

involves a new (complex) parameter s. It extends the density transformer since $\mathbf{G}_1 = \mathbf{G}$ and plays a crucial rôle in the analysis of rational trajectories. It acts on $A_\infty(D)$ as soon as $\Re s > 1/2$. Remark that its iterate $\mathbf{G}^n$ of order n involves the set $\mathcal{G}^n$ of the inverse branches of depth n,

$$\mathbf{G}_s^n[f](z) = \sum_{h\in\mathcal{G}^n} |h'(z)|^s \; f\circ h(z).$$

Example. Let us consider the Euclidean dynamical system related to the Gauss map to give a more precise view. Every real number $x \in]0,1]$ admits a continued fraction expansion of the form

$$x = \cfrac{1}{m_1 + \cfrac{1}{m_2 + \cfrac{1}{\ddots + \cfrac{1}{m_p + \dots}}}},$$

where the m_i form a sequence of positive integers. Ordinary continued fraction expansions of real numbers are the result of an iterative process which constitutes the continuous counterpart of the standard Euclidean division algorithm. They can be viewed as trajectories of a specific dynamical system relative to the Gauss map $T : [0,1] \to [0,1]$ defined by

$$T(x) := \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor, \quad \text{for } x \neq 0, \qquad T(0) = 0,$$

(here, $\lfloor x \rfloor$ is the integer part of x). The set $\mathcal{G}$ of the inverse branches of T is $\mathcal{G} := \{h : x \mapsto \frac{1}{m+x} \text{ for } m \geq 1\}$ and the transfer operator is

$$\mathbf{G}_s[f](z) = \sum_{h\in\mathcal{G}} |h'(z)|^s \; f\circ h(z) = \sum_{m\geq 1} \frac{1}{(m+z)^{2s}} \, f\left(\frac{1}{m+z}\right).$$

One of the main and most useful property of these transfer operators is that, under a "contracting condition," for real s the operator $\mathbf{G}_s$ has a unique dominant eigenvalue $\lambda(s)$, positive and isolated from the remainder of the spectrum by a spectral gap.

The constants we wish to compute are of three kinds and are all related to this dominant eigenvalue $\lambda(s)$;

1. Evaluate $\lambda(r)$ for some r real. For example $\lambda(2)$ is related to the coincidence probability and appears in the analysis of the height of digital trees.
2. Evaluate $\lambda'(r)$ or $\lambda''(r)$ for some r real. For instance, $-\lambda'(1)$ is the entropy of the dynamical system and plays a central rôle. The quantity $\lambda''(1)$ intervenes in the expression of the variance for the average number of steps of the classical Euclidean algorithm.
3. In the context of computing a local expansion of the quasi-inverse $(\mathrm{Id} - \mathbf{G}_s)^{-1}$, compute r such that $\lambda(r) = 1$. Also this is related to the computation of Hausdorff dimension of Cantor-like sets associated with incomplete dynamical systems.

2. The Principles of the Algorithm

For any function $f \in A_\infty(D)$, the Taylor expansions at $x_0 \in \mathcal{I}$ of f and $\mathbf{G}[f]$ exist and the operator $\mathbf{G}$ can be viewed as an infinite matrix $\mathbf{M} := (M_{i,j})$ with $0 \leq i, j < \infty$ and

$$M_{i,j} = \text{ the coefficient of } (z - x_0)^i \text{ in } \mathbf{G}[(Z - x_0)^j](z).$$

The truncated matrix $\mathbf{M}_n := (M_{i,j})_{0 \leq i,j \leq n}$ is the matrix of order $n+1$ which describes the action of a "truncated" operator on the space $\mathcal{P}_n$ formed with polynomials of degree at most n. More precisely, the truncated matrix $\mathbf{M}_n$ represents the truncated operator $\pi_n \circ \mathbf{G}|_{\mathcal{P}_n}$ where π_n is the projection on $\mathcal{P}_n$ which associates to a function f its Taylor expansion of order n at x_0 i.e.,

$$\pi_n[f](z) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} \, (z - x_0)^k.$$

Note that the operator $\pi_n \circ \mathbf{G}$ and the matrix $\mathbf{M}_n$ have the same spectrum.

In the case of the Euclidean dynamical system, Daudé, Flajolet, and Vallée introduced in [2] a method for computing (a finite part of) the spectrum of transfer operators, which they further used in [3, 8]. Their method, the so–called DFV–method, has three main steps which we describe in an informal way (see Figure 1).

(i) Compute the truncated matrix $\mathbf{M}_n$ relative to the operator $\mathbf{G}$.

(ii) Compute the spectrum $\operatorname{Sp} \mathbf{M}_n$ of matrix $\mathbf{M}_n$, i.e., the set of its eigenvalues,

$$\operatorname{Sp} \mathbf{M}_n := \{ \lambda_n^{(i)} : 0 \leq i \leq n \}.$$

(iii) Relate the set $\operatorname{Sp} \mathbf{M}_n$ with a (finite) part of $\operatorname{Sp} \mathbf{G}$.

$$\begin{array}{ccc} \text{Operator } \mathbf{G} & \xrightarrow{\text{Hard}} & \text{Eigenvalue } \lambda \\ {\scriptstyle \text{Truncate}}\downarrow & & \uparrow{\scriptstyle |\lambda_n - \lambda| \leq K\theta^n} \\ \text{Matrix } \mathbf{M}_n & \xrightarrow[\text{Easy}]{} & \text{Eigenvalue } \lambda_n \end{array}$$

FIGURE 1. The DFV-method for computing eigenvalue approximates.

In the case when the transfer operator has a unique dominant eigenvalue λ, isolated from the remainder of the spectrum, one can expect that it is the same for $\mathbf{M}_n$, with a dominant eigenvalue λ_n. Moreover, the authors of [2] observed that the sequence λ_n seems to converge to λ (when the truncation degree n tends to ∞), with exponential speed. They conjectured the following:

There exist n_0, K, θ such that, for any $n \geq n_0$, one has $|\lambda_n - \lambda| \leq K\theta^n$.

The main result here is that this can be proved in a very general setting, i.e., as soon as the dynamical system considered is strongly contracting and not just contracting.

A complete dynamical system is said to be *strongly contracting* if the set $\mathcal{G}$ of inverse branches fulfills the supplementary conditions: For any subset of the inverse branches $\mathcal{A} \subset \mathcal{G}$, there exist x_0 and two open disks of center x_0 D_S and D_L with $D_S \subsetneq D_L$ and $\mathcal{I} \subset D_L$, such that any $h \in \mathcal{A}$ is an element of $A_\infty(D_L)$ which strictly maps D_L inside D_S, i.e., $h(\overline{D}_L) \subset \overline{D}_S$.

Many dynamical systems relative to the Euclidean algorithms belong to this strongly contracting setting (and even to an extra-contracting setting where another disk comes between D_L and D_S and the constraint is put on the inverse branches of an iterate at a certain order of the map T – see [5] for a precise definition).

3. **Effective Computation**

One has under the strongly contracting condition that the truncated operator $\pi_n \circ \mathbf{G}$ converges in norm to $\mathbf{G}$. However an effective algorithm to solve the problem involves a circle $C = C(\lambda, r)$ (with center λ and radius $r > 0$) that isolates the dominant eigenvalue λ from the remainder of the spectrum and the quantity α_C defined by

$$\alpha_C(\mathbf{G}) := \sup_{z \in C} \|(\mathbf{G} - z\mathrm{Id})^{-1}\|.$$

In general it is an open problem to compute this quantity $\alpha_C(\mathbf{G})$ even if the dynamical system is strongly contracting. There is an exception when the operator satisfies a normality property on some functional space. Fortunately this is the case for the usual Euclidean dynamical system (but on an Hardy space and not $A_\infty(D)$). So this gives an effective algorithm to compute polynomially the constants. Finally evaluation of constants related to derivatives of the function $\lambda(s)$ is done thanks to Taylor expansion.

As a conclusion, let us just state the numerical value of the *Hensley constant* [4] with seven (proved) digits

$$\gamma_H = 2\frac{\lambda''(1) - \lambda'(1)}{\lambda'(1)^3} \approx 0.5160524\ldots$$

This constant appears in the variance of the average number of divisions P_n in the Euclid algorithm on a pair (u, v) such that $0 \leq u \leq v \leq n$ and which is $\mathbf{Var}\, P_n \sim \gamma_H \log n$. The reader is referred to [5] for numerical values of other constants like the *Gauss–Kuz'min–Wirsing constant* and the *Hausdorff dimension* of the Cantor set $\mathcal{R}_{\{1,2\}}$.

Bibliography

[1] Clément (Julien), Flajolet (Philippe), and Vallée (Brigitte). – Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica*, vol. 29, n° 1-2, 2001, pp. 307–369. – Average-case analysis of algorithms (Princeton, NJ, 1998).

[2] Daudé (Hervé), Flajolet (Philippe), and Vallée (Brigitte). – An average-case analysis of the Gaussian algorithm for lattice reduction. *Combinatorics, Probability and Computing*, vol. 6, 1997, pp. 397–433.

[3] Flajolet (Philippe) and Vallée (Brigitte). – Continued fraction algorithms, functional operators and structure constants. In *Proceedings of the Canadian Mathematical Society*, vol. Constructive, Experimental, and Nonlinear Analysis (in the honour of Jonathan Borwein). – 2000.

[4] Hensley (Doug). – The number of steps in the Euclidean algorithm. *Journal of Number Theory*, vol. 49, n° 2, 1994, pp. 142–182.

[5] Lhote (Loïck). – Computation of a class of continued fraction constants. In *Proceedings of the Sixth Workshop on Algorithm Engineering and Experiments and the First Workshop on Analytic Algorithmics and Combinatorics, New Orleans, LA, USA, January 10, 2004*, pp. 199–210. – 2004.

[6] Mayer (Dieter H.). – Spectral properties of certain composition operators arising in statistical mechanics. *Communications in Mathematical Physics*, vol. 68, 1979, pp. 1–8.

[7] Mayer (Dieter H.), T. Bedford (M. Keane) and Series (C.) (editors). – *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, Chapter Continued fractions and related transformations, pp. 175–222. – Oxford University Press, 1991.

[8] Vallée (Brigitte). – Dynamique des fractions continues à contraintes périodiques. *Journal of Number Theory*, vol. 72, n° 2, 1998, pp. 183–235.

[9] Vallée (Brigitte). – Dynamical sources in information theory: fundamental intervals and word prefixes. *Algorithmica*, vol. 29, n° 1-2, 2001, pp. 262–306. – Average-case analysis of algorithms (Princeton, NJ, 1998).

[10] Vallée (Brigitte). – Dynamical analysis of a class of Euclidean algorithms. *Theoretical Computer Science*, vol. 297, n° 1-3, 2003, pp. 447–486. – Latin American theoretical informatics (Punta del Este, 2000).

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 75–76.

Available online at the URL
http://algo.inria.fr/seminars/.

Profile of Random Recursive Trees and Random Binary Search Trees

Hsien-Kuei Hwang
Institute of Statistical Science, Academia Sinica (Taiwan)

April 26, 2004

Summary by Brigitte Chauvin and Jean-Maxime Labarbe

1. Introduction

Random recursive trees model is a simple probability model useful for many applications as system generation, spread of contamination of organisms, internet interface map, stochastic growth of networks or statistical physics. A random recursive tree is constructed as follows: one starts from a root node with the label 1; at each step $n, n \geq 2$, a new node with the label n is attached uniformly at random to one of the previous nodes labelled by $1, 2, \ldots, n-1$. In this model, the labels are increasing along any path from the root to a node.

Denote by $X_{n,k}$ the number of nodes at distance k from the root in a random recursive tree with n nodes. We are interested in the *profile* of random recursive trees, i.e., the collection $\{ X_{n,k} : k \in \mathbb{N} \}$. More precisely, there exist some results about the mean and the variance of the $X_{n,k}$ but the limit distribution of $X_{n,k}$ scaled by its mean, in the range $k/\log n \sim$ Constant is a source of intriguing phenomena. Notice that the profile provides a fine and informative shape characteristic, it is related to path length, depth, height, width, ... and also to generation of random trees or other algorithmic problems.

A well-known connection between random recursive trees and binary search trees is the following: by rotation, a planar tree gives a binary tree and then it appears that the profile of random recursive trees is exactly the "left" profile of binary search trees (meaning that the distance of a node is only counted for left branches). Consequently, it is of the same flavour to study the profile of binary search trees; results and phenomena for $Y_{n,k}$, the number of external nodes at level k and $Z_{n,k}$, the number of internal nodes at level k in a random binary search tree, appear as a simple transposition.

2. Main Results

Let k depend on n, let $\alpha_{n,k} = k/\log n$ and suppose that:

$$\lim_{n\to\infty} \alpha_{n,k} = \alpha$$

The main phenomena are summarized in the following proposition:

Proposition 1 (Main phenomena)**.**

- $\mathbf{E}(X_{n,k})$ *is unimodal and* $\mathbf{Var}(X_{n,k})$ *is bimodal,*
- $\forall \alpha \in [\,0, e)$, $X_{n,k}/\,\mathbf{E}(X_{n,k}) \xrightarrow{d} X_\alpha$ *(convergence in distribution)*
- $\forall \alpha \in [\,0, 1\,]$, $X_{n,k}/\,\mathbf{E}(X_{n,k}) \xrightarrow{m} X_\alpha$ *(convergence of all moments)*
- *for* $k = o(\log n)$*, (case* $\alpha = 0$*),* $(X_{n,k} - \mathbf{E}(X_{n,k}))/\sqrt{\mathbf{Var}(X_{n,k})} \xrightarrow{m} \mathcal{N}(0,1)$

– *for* $k = \log n + o(\log n)$ *(case* $\alpha = 1$*) and* $|k - \log n| \to \infty$,
$(X_{n,k} - \mathbf{E}(X_{n,k}))/\sqrt{\mathbf{Var}(X_{n,k})} \stackrel{m}{\to} X_1'$
– *for* $k = \log n + O(1)$, $(X_{n,k} - \mathbf{E}(X_{n,k}))/\sqrt{\mathbf{Var}(X_{n,k})}$ *does not converge in distribution.*

where X_α *and* X_1' *are limit distributions we describe further.*

The proof is based both on the contraction method and the moment method.

Let $\mu_{n,k}$ be the first moment of $X_{n,k}$. A fine study of the asymptotics of $\mu_{n,k}$ leads to

$$\text{for } \ 0 \le \alpha < e, \qquad \frac{\log \mu_{n,k}}{\log n} \longrightarrow \alpha - \alpha \log \alpha.$$

The second moment and the variance of $X_{n,k}$ can also be asymptotically described: if $0 \le \alpha < 2$,

$$\mathbf{Var}(X_{n,k}) \sim \Big(\frac{\Gamma(\alpha+1)^2}{(1-\alpha/2)\Gamma(2\alpha+1)} - 1 \Big) \mu_{n,k}^2$$

and the variance exhibits a *bimodal behavior* when $\alpha = 1$.

For the second point in the proposition, i.e., the limit distribution, the starting point is the recurrence formula satisfied by the $X_{n,k}$ (a branching-type property):

$$X_{n,k} \stackrel{d}{=} X_{U_n,k-1} + X^*_{n-U_n,k} \tag{1}$$

where U_n is uniform over $\{0, \dots, n-1\}$ and $X_{j,k}$ and $X^*_{j',k'}$ are independent of each other and independent of U_n. As usually, thanks to the asymptotics of the first moment $\mu_{n,k}$ of $X_{n,k}$ and because U_n/n converges in distribution to a uniform distribution U on the interval $[0,1]$, it is possible to deduce a limit equation from (1):

$$X_\alpha \stackrel{d}{=} \alpha U^\alpha X_\alpha + (1-U)^\alpha X^*_\alpha.$$

Moreover, the convergence of the first m moments is obtained for $0 \le \alpha < m^{1/(m-1)}$ and the moments of the limit distribution $\nu_m := \mathbf{E}(X_\alpha^m)$ are given by a recurrence relation:

$$\nu_m = \frac{1}{m - \alpha^{m-1}} \sum_{j=1}^{m} \binom{m}{j} \nu_j \nu_{m-j} \alpha^{j-1} \frac{\Gamma(j\alpha+1)\Gamma((m-j)\alpha+1)}{\Gamma(m\alpha+1)}.$$

In the particular case when $\alpha = 1$, the convergence of all moments in the fifth point of the Proposition is obtained by the method of moments: all moments satisfy the same type of recurrence:

$$a_{n,k} = b_{n,k} + \frac{1}{n-1} \sum_{j=1}^{n-1} (a_{j,k-1} + a_{j,k})$$

so that generating functions techniques and transfer theorem allow one to get asymptotics for higher moments. Notice that the limit distribution X_1' is nothing but $(dX_\alpha/d\alpha)|_{\alpha=1}$ and is a solution of a 'quicksort'-type equation:

$$X_1' \stackrel{d}{=} U X_1' + (1-U) X_1'^* + U + U \log U + (1-U)\log(1-U).$$

Among open questions:

– what happens at the boundary of the interval (α_-, α_+) of convergence of binary search trees (analog of the interval $(0, e)$ for recursive trees)?
– how to prove a.s. convergence in general?
– how to plot or simulate limit laws like X_α?

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 77–81.

Available online at the URL
http://algo.inria.fr/seminars/.

Airy Phenomena and the Number of Sparsely Connected Graphs

Bruno Salvy
Algorithms Project, Inria (France)

December 15, 2003

Summary by Vlady Ravelomanana

Abstract

The enumeration of connected graphs by excess (number of edges minus number of vertices) is a well-understood problem usually dealt by means of combinatorial decompositions or indirect formal series manipulations. In their work [3], Philippe Flajolet, Bruno Salvy, and Gilles Schaeffer derive such enumerative results mainly using analytic methods. In particular, they exhibit strong connections between Airy functions and the complete asymptotic expansions of the number of connected graphs of fixed excesses.

1. Introduction

It was Sir Edward M. Wright (1906–2005), of Hardy and Wright fame,[1] who initiated the enumeration of labelled connected graphs by number of vertices and edges [7, 8, 9]. The enumeration of graphs according to these two parameters has a long history which goes back to Cayley and whose main steps can be summarized as follows:

Author	Year	Results
Cayley	1889	number of unrooted trees
Rényi	1960	number of unicyclic graphs
Wright	1977	number of general connected graphs

These combinatorial problems are closely related to the theory of random graphs [2, 1, 4]. The starting point of Flajolet, Salvy and Schaeffer's work is the divergent series of connected labelled graphs

$$C(z,q) = \log\left(1 + \sum_{n=1}^{\infty} (1+q)^{\binom{n}{2}} \frac{z^n}{n!}\right). \tag{1}$$

(Throughout this abstract the variable z marks the number of vertices and the variable q reflects the number of edges.) Though Equation (1) can be viewed as an application of symbolic methods in combinatorial analysis [6], it is worthnoting that this series diverges for any $q > 0$. However, the paper [3] shows how to work with (1) mainly using analytical tools from asymptotic analysis. This abstract is divided into three parts as follows: (i) integral representation of (1), (ii) asymptotic expansions via standard saddle-point method and (iii) double saddle-point expansions and Airy functions.

[1] Sir Edward Wright was knighted in 1977 and a building in the University of Aberdeen beared in his very life. While his research career was striking long (1930–1980) and fruitful, he was also an excellent university admnistrator.

2. **Formal Expressions and Integral Representations**

Denote by $G(z,q)$ the bivariate EGF of labelled graphs (connected or not) then classically we obtain in $\mathbb{C}[[z,q]]$

$$G(z,q)=\sum_{n=0}^{\infty}(1+q)^{\binom{n}{2}}\frac{z^n}{n!}. \tag{2}$$

Thus, the EGF of connected graphs is given by

$$C(z,q)=\log\left(G(z,q)\right)=z+q\frac{z^2}{2!}+(3q^2+q^3)\frac{z^3}{3!}+(16q^3+15q^4+6q^5+q^6)\frac{z^4}{4!}+\cdots, \tag{3}$$

which is valid in $\mathbb{C}[[z,q]]$. Denote by $C_{n,n+\ell}$ the number of connected labelled graphs built with n nodes and $n+\ell$ edges. Let W_ℓ be the exponential generating function (EGF) of connected labelled graphs with ℓ edges more than vertices. Therefore, $W_\ell(z)=\sum_n C_{n,n+\ell}\frac{z^n}{n!}$ and in $\mathbb{C}[[z,q]]$ we have

$$\begin{aligned} Q(z,q) &:= \sum_{n,\ell} C_{n,n+\ell}(-q)^{\ell+1}\frac{z^n}{n!}=-qC(-z/q,-q) \\ &= W_{-1}(z)-qW_0(z)+q^2W_1(z)+\cdots=-q\log\left(\sum_{n=0}^{\infty}(1-q)^{\binom{n}{2}}\frac{(-zq^{-1})^n}{n!}\right). \end{aligned} \tag{4}$$

Observe that the *negatively signed* variable q represents an essential trick in the authors approach. In fact, the series (in z) $G(z,q)$ diverges as soon as $q>0$ is fixed but as pinpointed by the authors, this function can acquire *bona fide* analytic sense. In fact, if q is fixed and satisfies $0<q<2$ (so that $|1-q|<1$) by considering a weighting π that assigns to a graph g the weight $\pi(g):=(-q)^{(\#\text{edges}(g)-\#\text{vertices}(g))}$, Flajolet *et al.* introduced two analytic objects

$$\mathcal{H}(z,q):=\sum_{g\text{ graph}}\pi(g)\frac{z^{|g|}}{|g|!},\qquad \mathcal{Q}(z,q):=-q\sum_{g\text{ connected graph}}\pi(g)\frac{z^{|g|}}{|g|!}. \tag{5}$$

Another analytic ingredient derives from integral representations:

Lemma 1. *If $V(z)=\sum_n v_n z^n$ with $\sum_n|v_n|<\infty$ and if w is a real number s.t. $w\in(0,1)$ then*

$$\sum_n w^{n^2/2}v_n=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}V(e^{ix\sqrt{\log w^{-1}}})e^{-x^2/2}\,dx. \tag{6}$$

This Lemma is proved in [3, Lemma 1] by means of classical Fourier integral combined with interchange of summation and integration in infinite sequences. The main interest of (6) is that such tricky integral representation *linearizes* the *quadratic forms* present in the *exponents* of the concerned EGFs. Using this, the authors of [3] obtained from (5) that:

Lemma 2. *The generating function of connected graphs admits for $q\in(0,1)$ the integral representation*

$$\mathcal{Q}(z,q)=-q\log\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2}-z\frac{(1-q)^{-1/2}}{q}e^{ix\lambda(q)}\right)dx\right), \tag{7}$$

where

$$\lambda\equiv\lambda(q):=\sqrt{\log\frac{1}{1-q}}.$$

Now, interchange of limits and coefficients operators is obtained with the help of a result due to Gessel and Wang (who gave relations between inversions in rooted trees and connected graphs) combined with [3, Lemma 3 – p. 9]. More specifically, we have the following Lemma:

Lemma 3. *Assume that for each z with $|z| < 1/e$ and as $q \to 0^+$ the bivariate generating function $\mathcal{Q}(z,q)$ satisfies an asymptotic expansions,*

$$\mathcal{Q}(z,q) \sim \sum_{\ell \geq 0} \mathcal{W}_\ell(z)(-q)^\ell, \qquad |z| < e^{-1}, \qquad q \to 0^+, \tag{8}$$

for a sequence of functions $\mathcal{W}_\ell(z)$. Then, for each ℓ, the equality $\mathcal{W}_\ell(z) = W_\ell(z)$ holds where W_ℓ is given by $W_\ell(z) := \sum_n C_{n,n+\ell} \frac{z^n}{n!}$.

As a consequence, this allows to identify Q and $\mathcal{Q}$.

3. Single Saddle-Point Analysis and Asymptotic Expansions

The integral representation (7) serves as starting point to estimate $Q(z,q)$ as $q \to 0^+$. Recall the following results:

Theorem 1. *Let $t := T(z) = z \exp(T(z))$ where $T(z)$ is the generating function of rooted trees. Then, (i) $W_{-1}(z) = t - \frac{t^2}{2}$, (ii) $W_0(z) = \frac{1}{2}\log\frac{1}{1-t} - \frac{t}{2} - \frac{t^2}{4}$ and (iii) for $k \geq 1$, there exist polynomials A_k such that $W_k(z) = \frac{A_k(t)}{(1-t)^{3k}}$.*

The objective is to prove Theorem 1 by analysis of the integral representation (7) when t is restricted in some fixed interval $(0,a)$ with $a < 1$. In fact, when $q \to 0^+$ by setting $x\lambda = w$ the integrand rewrites as

$$\exp\left(-\frac{1}{q}\left(\frac{w^2}{2} + ze^{iw}\right)\right) \times \exp\left(\frac{w^2}{2}(1/q - 1/\lambda^2) + ze^{iw}\frac{1-(1-q)^{-1/2}}{q}\right) \tag{9}$$

and we see that the integrand can vary abruptly due to the factor $1/q$ above. The saddle-points of the first factor are located at points τ such that $\tau + ize^{i\tau} = 0$. Therefore, as $|z| < 1/e$, $\tau = -it = -iT(z)$. The saddle-point method consists in shifting the line of integration parallel to itself so that it crosses the point τ. Setting $w = u - it$ and replacing the contour on $(-\infty,\infty)$, we get

$$Q(z,q) \;=\; \left(t - \frac{t^2}{2}\right) + \left(1 - \frac{q}{\lambda^2}\right) - \frac{q}{\lambda\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{\left(\frac{u^2}{2} + t\left(e^{iu} - 1 - iu\right)\right)}{q}\right) h(u)\, du \tag{10}$$

where $h(u) = \exp\left((u^2/2 - uit)(q^{-1} - \lambda^{-2}) + te^{iu}(1-(1-q)^{-1/2})/q\right)$. Next, the kernel of the saddle-point integral is reduced to standard *quadratic form*. This is done using change of variable defined by $y^2 = f(u)$ with $f(u) := u^2/2 + t(e^{iu} - 1 - iu)$. This leads to another expression of the integral above, viz.

$$I = \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{q}\left(\frac{u^2}{2} + t\left(e^{iu} - 1 - iu\right)\right)\right) h(u)\, du = \int_{-\infty}^{+\infty} e^{-y^2/q} H(y)\, dy, \quad H(y) := h(u(y))\frac{du}{dy}. \tag{11}$$

Now, the expansion of H as a power series in y can be integrated term by term with the Gaussian kernel $\exp(-y^2/q)$ (the validity of this term by term integration is proved in [3]). Expansions at

any finite order with respect to q are legitimated and lead to finite order expansions (as $q \to 0^+$) as follows:

$$Q(z,q) \sim \left(T(z) - \frac{T(z)^2}{2}\right) - \left(\frac{1}{2}\log\frac{1}{1-T(z)} - \frac{T(z)}{2} - \frac{T^2(z)}{4}\right) q + \sum_{k\geq 2} \frac{A_{k-1}(T(z))}{(1-T(z))^{3k}}(-q)^k. \tag{12}$$

4. **Coalescing Saddle-Points and Airy Functions**

The aim of this section is to summarize the proof that the asymptotic number of connected graphs $C_{n,n+k}$ (for $k \geq 2$) can be expressed with $A_k(1)$ and the derivatives $A_k^{(j)}(1)$ ($j \geq 1$). Moreover, in their turn, the $A_k(1)$ (given also in [4]) can be expressed in terms of the Airy 'Ai' function [5]. Among other results, the authors of [3] obtained the following:

Theorem 2. *For any fixed value of k, the asymptotic number of connected graphs with n vertices and $n+k$ edges satisfied*

$$C_{n,n+k} = A_k(1)\sqrt{\pi}\left(\frac{n}{e}\right)^n \left(\frac{n}{2}\right)^{\frac{3k-1}{2}} \left(\frac{1}{\Gamma(3k/2)} + \frac{\frac{A_k'(1)}{A_k(1)} - k}{\Gamma(3k/2-1/2)}\sqrt{\frac{2}{n}} + O\left(\frac{1}{n}\right)\right). \tag{13}$$

The generating series of the dominant coefficients $A_k(1)$ has an asymptotic series

$$\sum_{k=1}^{\infty} A_k(1)(-x)^k = \log\left(1 + \sum_{k=1}^{\infty} c_k(-x)^k\right) \sim \log\left(2\sqrt{\pi i}(2x)^{-1/6} e^{1/(3x)}\,\mathrm{Ai}\left((2x)^{-2/3}\right)\right), \quad x \to 0, \tag{14}$$

where $c_k = (6k)!/((3k)!\,(2k)!\,3^{2k}2^{5k})$.

Observe first that the *single saddle-point* analysis of the previous section leads to an expansion valid for $t = T(z)$ in any closed interval included in $[0,1)$ and such an expansion becomes meaningless as t approaches 1. Thus, appropriate tools are needed. In fact as t approaches 1, *two coalescing saddle points* arise. The main steps of the analysis consist in (i) localizing the dominant saddle points, (ii) normalizing the integrand by means of a *cubic* change of variables, (iii) integrating formally term by term, (iv) analyzing the remander of the obtained expansions to legitimate the formal result, (v) reorganizing correctly such expansions.

After rescaling (by the change of variable $\alpha = q/\theta^3$ with $\theta = 1-t \equiv 1-T(z)$), the starting point is now the following integral representation:

$$I := \int_{-\infty}^{+\infty} e^{-f(u)/q} h(u)du, \text{ with } f(u) = \frac{u^2}{2} + (1-\theta)(e^{iu} - 1 - iu). \tag{15}$$

Note that $f(u)$ is now locally cubic near $\theta = 0$. Solving $f'(u) = 0$ for $u \neq 0$ reveals this time a (purely imaginary) saddle point $\rho = -2i\theta(1 + 1/3\theta + \ldots)$. The classical method of *coalescent saddle-points* is then useful to asymptotically estimate the integral defined in (15). Namely, it is convenient to introduce the cubic change of variable $f(u) = P(v)$ with $P(v) = f(\rho)/\theta^3(2v^3 + 3\theta v^2)$ where the polynomial P is s.t. P' has two roots at 0 and $-\theta$ and $P(0) = 0$, $P(-\theta) = f(\rho)$. Thus, P and f behave similarly in the neighborhood of their two central saddle points. The integral now admits the exact expression

$$I = -\int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{-P(v)/q} G(v)\,dv, \qquad G(v) = h(u(v))\frac{du}{dv}. \tag{16}$$

Then, like in the previous section the next step consists in expanding G as a power series in v: $G(v,\alpha,\theta)=\sum_k g_k(\alpha,\theta)v^k$ and integrating term by term. This process leads to

$$I \sim \sum_{k=0}^{\infty} g_k(\alpha,\theta)\theta^{k+1} R_k\left(\frac{-\theta^3}{f(\rho)}\alpha\right), \tag{17}$$

The exponential generating series of the $R_k(x)$ is given by

$$R(z) := \sum R_k(x)\frac{z^k}{k!} = \int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{1/x(2v^3+3v^2)+zv}\,dv. \tag{18}$$

Comparing this with the classical integral representation of the Airy function leads to

$$R(z) = 2\pi i\left(\frac{x}{6}\right)^{1/3}\exp\left(-\frac{z}{2}+\frac{1}{2x}-\frac{1}{2x}\left(1-\frac{2}{3}zx\right)^{3/2}\right)e^{\frac{2}{3}y^{3/2}}\,\mathrm{Ai}(y), \tag{19}$$

where $y=\left(1-\frac{2}{3}zx\right)\left(\frac{3}{4x}\right)^{2/3}$. After some analysis, it is shown that

$$R_k(x) \sim i\sqrt{\frac{\pi}{3}}c_k x^{\lfloor\frac{k+1}{2}\rfloor+1/2} \text{ as } x\to 0^+, \text{ where } c_{2k}=\frac{(-1)^k(2k)!}{12^k k!}, c_{2k+1}=\frac{(-1)^k(2k+3)!}{36(k+1)!12^k}.$$

The proof of [3, Theorem 2] is then completed by collecting the powers of θ in (17).

This short note is a summary of Ph. Flajolet, B. Salvy, and G. Schaeffer's article [3].

Bibliography

[1] Bollobás (Béla). – *Random graphs*. – Academic Press, 1985.

[2] Erdős (Paul) and Rényi (Alfred). – On random graphs. *Publicationes Mathematicae Debrecen*, vol. 6, 1959, pp. 290–297.

[3] Flajolet (Philippe), Salvy (Bruno), and Schaeffer (Gilles). – Airy phenomena and analytic combinatorics of connected graphs. *The Electronic Journal of Combinatorics*, vol. 11, n° R34, 2004. – Downloadable at `http://www.combinatorics.org/Volume_11/PDF/v11i1r34.pdf`.

[4] Janson (Svante), Knuth (Donald E.), Łuczak (Tomasz), and Pittel (Boris). – The birth of the giant component. *Random Structures Algorithms*, vol. 4, n° 3, 1993, pp. 231–358. – With an introduction by the editors.

[5] Whittaker (E. R.) and Watson (G. N.). – *A Course of Modern Analysis*. – Cambridge University Press, 1927. Reprinted 1973 (fourth edition).

[6] Wilf (Herbert). – *Generatingfunctionology*. – Academic Press, 1990. Avalaible freely online at `http://www.math.upenn.edu/~wilf/gfology.pdf`.

[7] Wright (Edward Maitland). – The number of connected sparsely edged graphs. *Journal of Graph Theory*, vol. 1, 1977, pp. 317–330.

[8] Wright (Edward Maitland). – The number of connected sparsely edged graphs ii: Smooth graphs and blocks. *Journal of Graph Theory*, vol. 2, 1978, pp. 299–305.

[9] Wright (Edward Maitland). – The number of connected sparsely edged graphs iii: Asymptotic results. *Journal of Graph Theory*, vol. 4, 1980, pp. 393–407.

Part III

Analysis of Algorithms and Protocols

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 85–92.

Available online at the URL
http://algo.inria.fr/seminars/.

On the Complexity of a Gröbner Basis Algorithm

Magali Bardet
Spaces Project, LIP6 and INRIA (France)

November 25, 2002

Summary by Bruno Salvy

Abstract

While the computation of Gröbner bases is known to be an EXPSPACE-complete problem, the generic behaviour of algorithms for their computation is much better. We study generic properties of Gröbner bases and analyse precisely the best algorithm currently known, F_5.

1. Gröbner Bases

Gröbner bases are a fundamental tool in computational algebra. They provide a multivariate generalization of Euclidean division and Euclid's algorithm for the gcd, as well as a generalization of Gaussian elimination to higher degrees. A very clear introduction is given in [3]; in this section we recall the basic definitions and properties.

1.1. **Definitions.** We consider polynomials in $\mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_n]$, where $\mathbb{K}$ is a field. The first step is to define a generalization of the univariate degree.

Definition 1. A *monomial ordering* is a total order on the set of monomials $\mathbf{x}^\alpha$ that is compatible with the product and such that 1 is the smallest monomial.

A monomial ordering can be given by a nonsingular real matrix A: the vectors of exponents are multiplied by the matrix and the resulting vectors are compared lexicographically. A technical condition encodes that 1 is minimal. Basic examples of orderings are: the lexicographic order, with identity matrix; the total degree order, also called *grevlex*, whose matrix has a first line of 1's above an antidiagonal matrix of -1's; the elimination orders whose matrix decompose diagonally into blocks of grevlex matrices. An order is said to *refine the degree* when the corresponding matrix has a fist line of 1's.

A polynomial can be expanded as a sum of terms, each term being a monomial times a coefficient. The *leading term* $\mathrm{LT}(p)$ of a polynomial p is then defined with respect to any monomial ordering.

The next step is to define an analogue of the Euclidean division. This process is called *reduction*, it depends on a given monomial ordering. Given a polynomial f and a set of polynomials $B = \{f_1, \ldots, f_s\}$, it returns a polynomial r such that

$$f = a_1 f_1 + \cdots + a_s f_s + r, \qquad \text{where } a_i \in \mathbb{K}[\mathbf{x}] \text{ for } i = 1, \ldots, s,$$

and the leading monomials of the f_i's do not divide that of r. One says that f *reduces* to r by B.

Definition 2. A *Gröbner basis* of an ideal $\mathfrak{I} \subset \mathbb{K}[\mathbf{x}]$ for a given monomial ordering is a finite set $B \subset \mathfrak{I}$ such that any $f \in \mathfrak{I}$ reduces to 0 by B. The basis is called *reduced* when the f_i's all have leading coefficient 1 and when none of the f_i's involves a monomial which reduces by $B \setminus \{f_i\}$.

An important consequence of the Hilbert basis theorem is the existence of Gröbner bases, thus of finite sets of generators, for all polynomial ideals. For a given monomial ordering, any ideal has a single reduced Gröbner basis.

1.2. **Examples.**

Example 1: gcd. In the univariate case, $\mathbb{K}[x]$ is principal, meaning that any ideal can be generated by a single generator. One possible choice for the generator is the gcd of its elements, which is also the only element of its reduced Gröbner basis.

Example 2: an intersection. Consider the system $\{f = x^2 + y^2 - 2, g = xy - 1\}$. These equations describe the intersection of a circle and a hyperbola, at points where they are tangent. The Gröbner basis for the ideal generated by $\{f, g\}$ for the lexicographic order with $x \prec y$ is $\{f_1 = (y^2 - 1)^2, f_2 = x - 2y + y^3\}$, while the Gröbner basis for the total degree order is $\{f, g, f_2\}$. Note that the multiplicities are preserved in this computation.

1.3. **Applications.**

Polynomial-system solving. Like in the previous example, using a lexicographic order yields a triangular system that can then be solved equation by equation.

Elimination. In Example 2, f_1 is a polynomial where the variable x has been eliminated between f and g. More generally, elimination can be computed using elimination orders. Geometrically, elimination corresponds to projection. It can be used to compute implicitizations, envelopes,....

Nullstellensatz. This answers the question: does p vanish on the common roots of $(f_1, \ldots, f_s)$? For instance, the polynomial $x - y$ vanishes on the common roots of f and g in our example. This is determined by computing a Gröbner basis of $(f_1, \ldots, f_s, 1 - tp)$ for a new variable t and observing that the result is $\{1\}$ (for any order).

Ideal membership. This answers the question: does p belong to the ideal generated by $(f_1, \ldots, f_s)$? This is decided by reducing p by a Gröbner basis and checking whether the result is 0 or not. Note that in our example, $x - y$ does not belong to the ideal, but $(x - y)^2$ does.

2. Worst-Case Complexities

The worst-case complexity of Gröbner bases has been the object of extensive studies. We refer to [8] for a survey.

2.1. **Polynomial-system solving is hard.** Since Gröbner bases can be used to solve polynomial systems, their complexity is at least that of polynomial-system solving. It turns out that it is not difficult to encode NP-complete problems into polynomial systems, which shows that the worst-case complexity cannot be expected to be too good. We give two examples.

Knapsack problem. Given $n+1$ natural integers $(b_1, \ldots, b_n, c)$, the problem of solving the overdetermined system

$$\sum_{i=1}^{n} x_i b_i = c, \qquad x_i(1 - x_i) = 0, \quad i = 1, \ldots, n$$

is known as the 0-1 knapsack problem and has been proved to be NP-complete by Karp in 1972.

3-SAT. Given Boolean variables X_i and a number of Boolean clauses each with three literals, i.e., clauses of the form

$$Y_j \vee Y_k \vee Y_\ell, \qquad (Y_j, Y_k, Y_\ell) \in \{X_1, \ldots, X_n, \neg X_1, \ldots, \neg X_n\},$$

3-SAT is the problem of deciding whether there exists a Boolean assignment to the X_i's that makes all the clauses true simultaneously (SAT stands for satisfiability). This is cast into an overdetermined polynomial system using the correspondence $X_i \mapsto x_i$, $\neg X_i \mapsto 1 - x_i$, $X \vee Y \mapsto x + y - xy$, together with the equations $x_i(1 - x_i) = 0$. 3-SAT has been proved to be NP-complete by Cook in 1971.

2.2. **From bad to worse.** Another problem solved by Gröbner bases turns out to have a much worse complexity: Ideal membership is EXPSPACE-complete. This means that any problem that can be solved with exponential space can be reduced to Ideal membership. We recall that complexity classes are ordered as follows:

$$\text{P} \subset \text{NP} \subset \text{PSPACE} \subset \text{EXPTIME} \subset \text{EXPSPACE}.$$

One source of this difficulty comes from multiplicities. Indeed, the Nullstellensatz problem is "only" in PSPACE. Another progress is made if one restricts attention to polynomial systems with only finitely many solutions (these are called *0-dimensional*). The computation of their Gröbner bases is also in PSPACE. If one furthermore demands that after homogenizing the polynomials the system still has finitely many (projective) solutions, then the computation of Gröbner bases falls into NP.

For s equations of degree at most d in n variables, the arithmetic complexity bounds for Gröbner bases are $2^{2^{O(n)}}$ in general, $d^{O(n^2)}$ in the 0-dimensional case and $s^{O(1)}d^{O(n)}$ when the homogenized system has finitely many solutions. These bounds should be compared with Bézout's theorem, stating that the number of solutions, when finite, is bounded by d^n, and is exactly d^n in the homogeneous case.

This picture leads to natural questions that are (partially) adressed in this work:

> *Where are "random" systems? What is the exponent hidden in their $O()$ term? What about* overdetermined *systems having solutions?*

3. Generic Systems and the F_5 Algorithm

We do not deal directly with random systems, but rather with generic ones. We now briefly recall what *generic* means in an algebraic context and describe the generic behaviour of the F_5 algorithm, of which we introduce a simple matrix version.

3.1. Genericity.

Definition 3. A property of points in a space of dimension N is *generic* when it holds at all points except on an algebraic set of dimension at most $N - 1$. (Here, an algebraic set is defined as the zero set of a system of polynomials).

Example. Two univariate polynomials $A = a_0x^{d_1} + \cdots + a_{d_1}$ and $B = b_0x^{d_2} + \cdots + b_{d_2}$ of degree d_1 and d_2 are generically relatively prime. Indeed, the pair (A, B) can be viewed as a point in a space of dimension $d_1 + d_2 + 2$, with coordinates the a_i's and b_i's. Their gcd is one if and only if there does not exist nonzero polynomials u and v with $\deg u < d_2$ and $\deg v < d_1$ such that $uA + vB = 0$. This is a linear system in the coefficients of u and v that has nonzero solutions if and only if the

determinant of the following *Sylvester matrix* is 0:

$$\begin{pmatrix} a_0 & a_1 & \dots & & \\ & \ddots & & \ddots & \\ & & a_0 & \dots & \\ b_0 & b_1 & \dots & & \\ & \ddots & & \ddots & \\ & & b_0 & \dots & \end{pmatrix}.$$

This determinant is a polynomial in the coordinates of (A, B) (the *resultant* of A and B), which shows that the "bad" points belong to an algebraic set. In order to prove that this algebraic set had dimension smaller than that of the space, it is sufficient to exhibit one point outside of it. Thus the proof is concluded by observing that $X^{d_1} \wedge (X^{d_2} + 1) = 1$.

When the base field $\mathbb{K}$ is $\mathbb{C}$ or $\mathbb{R}$, generic properties hold outside a set of measure 0. When $\mathbb{K}$ is $\mathbb{Q}$ or a finite field with large enough characteristic, then quantitative probability bounds can be obtained in terms of the *degree* d of the algebraic set. For any $S \subset \mathbb{K}$, a point whose coordinates are chosen independently with uniform probability from S has probability at least $1 - d/|S|$ to lie outside of the algebraic set [9, 11]. Thus "generic" is related to "random" in a very precise way.

3.2. **Buchberger's algorithm.** In view of our definition of Gröbner bases above, a property (which could be taken as a definition) is that each element of the ideal has a leading monomial which is a multiple of that of one of the elements of the basis. Buchberger's algorithm consists in producing repeatedly new leading monomials using S-polynomials.

Definition 4. Let f and g be two polynomials and m be the lcm of their leading monomials, then the *S-polynomial* of f and g is

$$S(f, g) := \frac{m}{\mathrm{LT}(f)} f - \frac{m}{\mathrm{LT}(g)} g.$$

In the univariate case, $S(f, g)$ corresponds to the first step in the Euclidean division of f by g. Buchberger's algorithm then proceeds as follows:

Initialization: $B := \{\}$, $S := \{f_1, \dots, f_s\}$
while $S \neq \{\}$ **do**
– pick $f \in S$; $S := S \setminus \{f\}$; reduce f w.r.t. B and call g the resulting polynomial;
– **if** $g \neq 0$ **then** $S := S \cup_{b \in B} S(g, b)$; add g to B
return B

Buchberger proved in his thesis (in 1965) that this algorithm terminates and produces a Gröbner basis. One of the main difficulties with an actual implementation is that the reduction steps often produce 0 and a lot of time is wasted during these useless reductions. Thus, there are many strategies to help "pick" an element in S and predict useless reductions.

3.3. **Macaulay's matrix.** Another approach to polynomial-system solving was described by Macaulay in [7] where he generalized Sylvester's matrix to multivariate polynomials. The idea is to construct a matrix whose lines contain the multiples of the polynomials in the original system, the columns representing a basis of monomials up to a given degree. It was observed by Lazard [6] that for a large enough degree, ordering the columns according to a monomial ordering and performing row reduction without column pivoting on the matrix is equivalent to Buchberger's algorithm. In this correspondence, reductions to 0 correspond to lines that are linearly dependent upon the previous ones and the leading term of a polynomial is given by the leftmost nonzero entry in the corresponding line.

3.4. F_5 **algorithm.** From now on and except in the last section, we restrict attention to fields of coefficients with characteristic 0 and homogeneous polynomials. Given a system of polynomials $f_1, \ldots, f_s$, with $\deg f_i =: d_i$ and $d_1 \leq \cdots \leq d_s$, we denote by $\mathcal{F}_i$ the sub-system $f_1, \ldots, f_i$, by $I(\mathcal{F}_i)$ the ideal it generates and by $I_d(\mathcal{F}_i)$ the vector space of elements of $I(\mathcal{F}_i)$ with degree d.

Faugère's F_5 algorithm [5] avoids "useless" lines coming from the relations $f_i f_j = f_j f_i$. We now present a matrix version of this algorithm. The algorithm is incremental in d, then in i. It constructs submatrices $M_{d,i}$ of the Macaulay matrix and performs a row reduction on them. The incremental step from $i-1$ to i introduces the lines corresponding to mf_i for all monomials m of degree $D - d_i$ that *do not appear as leading monomials* in the reduced $M_{D-d_i,i-1}$. This matrix is then reduced and stored in $M_{d,i}$. The algorithm stops when a large enough D has been reached.

The number of linearly independent lines in the matrix $M_{d,s}$ is the number of linearly independent polynomials in $I_d(\mathcal{F}_s)$. Subtracting this from the number of monomials of degree d (the number of columns of the matrix), one gets a function $\mathrm{HF}(d)$ known as the *Hilbert function* of the ideal. For large enough d, this function is a polynomial in d (the *Hilbert polynomial*). The generating series $H(z) = \sum_{d \geq 0} \mathrm{HF}(d) z^d$ is called the *Hilbert series* and geometric information related to the algebraic set can be read off from it. The smallest value of d such that the Hilbert function is equal to the Hilbert polynomial is called the *index of regularity* $i_{\mathrm{reg}}(I)$ of the ideal. The homogeneity hypothesis makes the above quantities intrinsic to the ideal, that is, they do not depend on the chosen ordering.

3.5. **Regular systems.** A striking result of [5] is that for *regular systems*, F_5 does not perform any useless reduction to 0.

Geometrically, the system $\mathcal{F}_s$ is regular when for each $i = 1, \ldots, s$, the algebraic set defined by $\mathcal{F}_i$ has codimension i. Algebraically, this is expressed by the fact that f_i is not a zero-divisor in the quotient $\mathbb{A}_i := \mathbb{K}[\mathbf{x}]/I(\mathcal{F}_{i-1})$. In other words, if there exists g such that $gf_i = 0$ in $\mathbb{A}_i$, then $g \in I(\mathcal{F}_{i-1})$. It is not difficult to see that among systems of degrees $(d_1, \ldots, d_s)$, the regular ones are generic. Classical properties of regular systems are: *(i)* the system $\mathcal{F}_s$ is regular if and only if its Hilbert series is given by

$$H(z) = \frac{\prod_{i=1}^{s} (1 - z^{d_i})}{(1-z)^n}; \tag{1}$$

(ii) the index of regularity is

$$1 + \sum_{i=1}^{s} (d_i - 1); \tag{2}$$

(iii) after a generic linear change of variables, the highest degree of elements of a Gröbner basis for the *grevlex* order is the index of regularity.

3.6. **Semi-regular systems.** Regular systems have at most as many polynomials as variables. We now generalize this definition, before stating our main result on the complexity of F_5.

Definition 5. A 0-dimensional system $\mathcal{F}_s$ is *semi-regular* when $gf_i = 0$ in $\mathbb{A}_i$ and $\deg(gf_i) < i_{\mathrm{reg}}(I(\mathcal{F}_s))$ imply $g \in I(\mathcal{F}_{i-1})$, for $i = 1, \ldots, s$.

The system $\mathcal{F}_s$ is semi-regular if and only if its Hilbert series is $[H(z)]$. Here, the bracket of a power series f is a power series whose coefficients are 0 starting at the index of the first negative coefficient of f, and are those of f before. It follows from this series that 0-dimensional regular systems are semi-regular; this new definition also accommodates overdetermined systems. The following proposition gives a way to compute i_{reg} efficiently.

Proposition 1. *For a semi-regular system, the degree of regularity is the index of the first non-positive coefficient in the series* (1).

We are now in a position to state the main result of this work:

Theorem 1. [1] *For a semi-regular system,* (i) *there is no reduction to 0 in the algorithm F_5 for degrees smaller than i_{reg};* (ii) *the number of operations in $\mathbb{K}$ performed by F_5 is bounded by*

$$O\left(\binom{i_{reg}+n}{n}^{\omega}\right).$$

The exponent ω is the exponent in the complexity of matrix multiplication. The best known bound for general matrices in characteristic 0 is $\omega < 2.39$. We refer to [2, Chapters 15–16] for these questions.

4. Asymptotic Analysis

If $i_{\mathrm{reg}} \sim \lambda n$ as $n \to \infty$, then the logarithm of the binomial in Theorem 1 is equivalent to $((1+\lambda)\ln(1+\lambda) - \lambda \ln \lambda)n$, while a "natural" size of the problem given by Bézout's theorem is $n \ln d$. We now describe how precise asymptotic information on i_{reg} can be obtained for semi-regular systems. Since the case when $s \leq n$ is given by (2), we concentrate on the overdetermined case.

4.1. **Principle.** The pth coefficient of the series (1) is given by the Cauchy integral representation

$$C(p) = \frac{1}{2i\pi} \oint \frac{\prod_{i=1}^{s}(1-z^{d_i})}{(1-z)^n} \frac{dz}{z^{p+1}}. \tag{3}$$

A preliminary analysis reveals that the degree of regularity grows roughly linearly with n. The analysis is then based on computing the asymptotic expansion of $C(\lambda n)$ for fixed λ, and then determining an asymptotic expansion $\lambda(n)$ that makes this behaviour vanish asymptotically.

4.2. **Few more equations than unknowns.** When $s = n + k$, with fixed k, it is convenient to rewrite the integral (3) as

$$C(p) = \frac{1}{2i\pi} \oint \underbrace{\prod_{i=1}^{s} \frac{1-z^{d_i}}{1-z} \frac{1}{z^{p+1}}}_{F_p(z)} (1-z)^k \, dz.$$

The coefficients can then be analyzed precisely using the *saddle-point method*. The integral is concentrated in the neighborhood of a saddle point ρ, characterized by $F_p'(\rho) = 0$. In the neighborhood of this point, the integrand behaves like $\exp(cz^2)$, and the next step of the method is to perform the quadratic change of variables $F_p(z) = F_p(\rho)\exp(-u^2)$. The integral is then approximated by

$$\frac{F_p(\rho)}{2i\pi} \int_{-\infty}^{\infty} e^{-u^2} (1-z(u))^k \frac{dz}{du} \, du. \tag{4}$$

The value of i_{reg} is obtained by choosing p such that this integral vanishes. At the first order, this is achieved by taking p such that $z(u) \sim \rho = 1$. Injecting this estimate in $F_p'(\rho) = 0$ gives the dominant term of the behaviour. The next one is obtained by renormalizing (4) in terms of the kth *Hermite polynomial* that satisfies

$$H_k(x) = \frac{2^k}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} (x+iu)^k \, du.$$

The final result is the following.

Theorem 2. *The degree of regularity of a semi-regular system of $s = n+k$ homogeneous polynomials of degree $d_1, \dots, d_{n+k}$ in n variables behaves asymptotically like*

$$\sum_{i=1}^{s} \frac{d_i - 1}{2} - \alpha_k \sqrt{\sum_{i=1}^{s} \frac{d_i^2 - 1}{6}} + O(1), \qquad n \to \infty,$$

where α_k is the largest zero of the kth Hermite polynomial.

It is actually possible to compute a full asymptotic expansion. For $s = n + 1$, $\alpha_1 = 0$ and the result found is in agreement with the exact result due to Szanto [10]. Also, from the practical point of view, this result shows in particular that linear equations do not increase the complexity.

4.3. **More equations.** The theorem above quantifies the gain in complexity obtained by adding more information in the form of extra equations. We now perform a similar analysis for systems with αn equations, $\alpha \geq 1$ being fixed.

In this case, the factor $(1 - z)^k$ is not a small perturbation any longer. The behaviour of the integrand changes qualitatively and the integral is then dominated by *two conjugate saddle points* $R_\pm$. The contributions of these saddle points to the integral are conjugate values whose sum does not vanish. This qualitative analysis reveals that a new phenomenon must occur for the integral to vanish: the index p must be such that *the saddle points coalesce*, giving rise to a double saddle point. This happens when both F' and F'' vanish and these equations are sufficient to give the first-order behaviour of i_{reg}, where now

$$F = \frac{\prod_{i=1}^{s} (1 - z^{d_i})}{(1 - z)^n} \frac{1}{z^{p+1}}.$$

A more precise analysis is achieved by capturing the coalescence of R_+ and R_- by means of a cubic change of variables $F(z) = P(u) = \exp(\frac{u^3}{3} + au^2 + c)$, where a and c are chosen so that the values of P at its saddle points 0 and $-2a$ is the same as that of F at R_- and R_+. The integral is then renormalized to

$$\frac{1}{2\pi} \int \exp P(u)\, du = \exp(c + \frac{2}{3}a^3)\, \mathrm{Ai}(a^2),$$

where Ai is the classical Airy function. The technicalities omitted here lead to the following result.

Theorem 3. *The degree of regularity of a semi-regular system of $s = \alpha n$ homogeneous polynomials of degree $d_1, \dots, d_{\alpha n}$ in n variables behaves asymptotically like*

$$\phi(\rho) n - a_1 \left(\frac{9}{2} \rho^2 \phi''(\rho) \right)^{1/3} n^{1/3} + \cdots, \qquad n \to \infty$$

where

$$\phi(z) = \frac{z}{1 - z} - \frac{1}{n} \sum_{i=1}^{s} \frac{d_i z^{d_i}}{1 - z^{d_i}},$$

ρ is the positive zero of $\phi'(z)$, and a_1 is the largest zero of the Airy function.

Moreover, in the neighbourhood of $\alpha = 1$, one gets

$$\phi(\rho) = \frac{1}{n} \sum_{i=1}^{s} \frac{d_i - 1}{2} - \sqrt{\sum_{i=1}^{s} \frac{d_i^2 - 1}{3n}} \sqrt{\alpha - 1} + \cdots$$

which is consistent with our previous result.

5. Extensions

5.1. **Affine case.** Up to now, we have considered only systems of homogeneous polynomials. When given nonhomogeneous polynomials, it is always possible to use an extra variable x_0 to make them homogeneous, choose a monomial order that makes this variable x_0 smaller than the other ones, compute the corresponding Gröbner basis, and set x_0 to 1 in the result. This gives the correct Gröbner basis and some of our analysis applies. However, in the overdetermined case, the homogenized system is not semi-regular (it is not 0-dimensional). It is therefore necessary to refine the analysis. This is done in [1].

5.2. **Positive characteristic.** An important application of Gröbner bases in cryptography involves overdertermined systems over the field $\mathbb{F}_2$ with two elements and moreover the solutions themselves are sought in $\mathbb{F}_2$. In that case, it is convenient to modify the algorithm F_5 so that "useless" lines coming from $f_i^2 = f_i$ are not computed. This results in an efficient algorithm that has been used to break a cryptographic challenge [4]. The analysis proceeds as before, the degree of regularity being now the first nonpositive coefficient in the series

$$\frac{(1+z)^n}{\prod_{i=1}^{s}(1+z^{d_i})}.$$

Bibliography

[1] Bardet (Magali), Faugère (Jean-Charles), and Salvy (Bruno). – On the complexity of Gröbner basis computation for regular and semi-regular systems. – 2005. In preparation.

[2] Bürgisser (Peter), Clausen (Michael), and Shokrollahi (M. Amin). – *Algebraic complexity theory*. – Springer-Verlag, Berlin, 1997, *Grundlehren der Mathematischen Wissenschaften*, vol. 315, xxiv+618p.

[3] Cox (David), Little (John), and O'Shea (Donal). – *Ideals, varieties, and algorithms*. – Springer-Verlag, New York, 1997, second edition, xiv+536p.

[4] Faugère (J.-C.) and Joux (A.). – Algebraic cryptanalysis of hidden field equation (HFE) cryptosystems using gröbner bases. In Boneh (D.) (editor), *Crypto'2003. Lecture Notes in Computer Science*, pp. 44–60. – Springer-Verlag, 2003.

[5] Faugère (Jean-Charles). – A new efficient algorithm for computing Gröbner bases without reduction to zero (F_5). In Mora (Teo) (editor), *ISSAC 2002*. pp. 75–83. – ACM Press, 2002. Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation, July 07–10, 2002, Université de Lille, France.

[6] Lazard (D.). – Gröbner bases, Gaussian elimination and resolution of systems of algebraic equations. In *Computer algebra*, pp. 146–156. – Springer, Berlin, 1983. Proceedings EUROCAL'83, London 1983.

[7] Macaulay (F. S.). – *The algebraic theory of modular systems*. – Cambridge University Press, Cambridge, 1994, *Cambridge Mathematical Library*, xxxii+112p. Revised reprint of the 1916 original.

[8] Mayr (Ernst W.). – Some complexity results for polynomial ideals. *Journal of Complexity*, vol. 13, n° 3, 1997, pp. 303–325.

[9] Schwartz (J. T.). – Fast probabilistic algorithms for verification of polynomial identities. *Journal of the Association for Computing Machinery*, vol. 27, n° 4, 1980, pp. 701–717.

[10] Szanto (Agnes). – Multivariate subresultants using Jouanolou's resultant matrices. *Journal of Pure and Applied Algebra*, 2004. – To appear.

[11] Zippel (Richard). – Probabilistic algorithms for sparse polynomials. In *Symbolic and algebraic computation (EUROSAM '79, Internat. Sympos., Marseille, 1979)*, pp. 216–226. – Springer, Berlin, 1979.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 93–96.

Available online at the URL
http://algo.inria.fr/seminars/.

A Probabilistic Counting Algorithm

Marianne Durand
Algorithms Project, INRIA (France)

December 9, 2002

Summary by Pierre Nicodème

Abstract

This talk[1] (a joint work with Philippe Flajolet) presents an algorithm to approximate count the number of different words in very large sets or texts (in the range of billions of bytes) and its analysis. When using an auxiliary memory of m bytes, the accuracy is of the order $1/\sqrt{m}$. The analysis of this new algorithm relies on asymptotic Depoissonization techniques.

1. Introduction

The problem addressed by this work is how to estimate the number of distinct elements in a large collection of data with the following requirements while: doing a single pass on the data; using a small amount of memory; doing only a few computations; doing no assumptions about the distribution of the data.

Applications of such a problem are data mining optimizations and routers programmation.

2. Summary of Some Algorithms

2.1. **Previous work.** In the following, the input words are considered as elements of $[0, 1]$ (take the 2-adic value of each word represented as a bit-string). We consider the number of input items N, the standard deviation σ of the rate of error done on the counting and the space S needed by the algorithm. In any case a hashing of the data is done before performing the algorithms. The algorithms precedently used for these aims may be classified as:

- **(adaptive) hashing schemes** [3]
  - algorithm: hash the values in $[0, 1/2]$ in a table; when the number of collisions exceeds a given value γ, skip to a different and smaller interval (by instance $[1/2, 3/4]$); return a function of the number of collisions.
  - parameters: $\sigma = 1.5/\sqrt{S}$; assumptions on the size of the data; unstable with respect to the order of arrival of the data;
- **adaptive sampling** [4]
  - algorithm: maintain one bucket of size m; when the bucket overflows, throw away all data beginning by a 1; filter out incoming data beginning by a 1; repeat the process by filtering with 00 and so on; return a function of the number of iterations;
  - parameters: $\sigma = 6.7/\sqrt{m}$ (unprecise algorithm);
- **probabilistic counting** [6]

[1] The results presented here and recent improvements will be presented at the ESA 2003 Symposium [2].

- ∗ algorithm: let $\rho(w)$ be the position of the first 1 in w; (by instance, $\rho(00010111) = 4$); set up the bits corresponding to $\rho(w)$ for all w in a bit-map. Let k_{max} be the first bit equal to 0 in this bit-map. The estimator is $2^{ck_{max}}$ for a given c.
- ∗ parameters: $\sigma = 0.78/\sqrt{m}$, $S = m \times \log_2 N$;

See also [1].

2.2. **The new algorithm of Durand and Flajolet.** This algorithm uses a technique of *maximum-based probabilistic counting*. It has the following features:

– algorithm: send the data to $m = 2^b$ different buckets, according to the value of their b first bits. For each bucket i compute the maximum

$$M^{(i)} = \max(\{\rho(\mathrm{suf}(w)); w \text{ is hashed in bucket } i\}),$$

where $\mathrm{suf}(w)$ is the suffixe of w starting at position $b + 1$.

$$\text{Return} \quad E = \alpha_m m \times 2^{\frac{1}{m}\sum M^{(i)}}, \tag{1}$$

$$\text{where} \quad \alpha_m = \left(\Gamma(-1/m)\frac{1 - 2^{1/m}}{\log 2}\right)^{-m}, \qquad \Gamma(s) = \frac{1}{s}\int_0^\infty e^{-t}t^s dt; \tag{2}$$

– parameters: $\sigma = 1.3/\sqrt{m}$; memory $S = m \times \log\log \max(\{M^{(i)}\})$;
– remarks: the algorithm is independent of the repetitions and need very few computations. It is only necessary to maintain one value of size $O(\log\log N)$ for each bucket and not a bitmap of size $O(\log N)$ as in probabilistic counting.

3. **Analysis**

As frequently observed, the analysis is easier when Poissonization-Depoissonization is used. The steps of the analysis therefore are.

1. Compute the generating function $F(z) = \sum f_n z^n$, where f_n is the estimator of number of different items when exactly n are read by the algorithm.
2. "Poissonize" the system by considering that the number of items read by the algorithm is a random number following a Poisson distribution of parameter λ. Asymptotically, when $\lambda = n$, one expects the Poisson model to reflect corresponding properties of the fixed-n model (note that for large λ, the Poisson law is concentrated near its mean).
   During this step, compute
   $$\widetilde{F}(z) = \sum f_n \frac{z^n}{n!} e^{-z} = \sum \widetilde{F}_n z^n.$$
3. Compute the Mellin transform $f^\star(s)$ of $\widetilde{F}(z)$. The expansions of $f^\star(s)$ in the neighborhood of its singularities give the asymptotic value of $\widetilde{F}_n$.
4. Prove by depoissonization that, asymptotically, $f_n \sim \widetilde{F}_n$.

3.1. **Getting the basic generating function.** We are interested here to the statistics of the estimator

$$Z = E/\alpha_m = m \times 2^{\frac{1}{m}\sum_i M^{(i)}}.$$

Considering one bucket that receives ν elements, the random variable M is the maximum of ν random variables Y that are independent and geometrically distributed according to $\mathbf{P}(Y \geq k) = 1/2^k$.

Therefore we have

$$\mathbf{P}_\nu(M \le k) = \left(1 - \frac{1}{2^k}\right)^\nu, \quad \text{and} \quad \mathbf{P}_\nu(M = k) = \left(1 - \frac{1}{2^k}\right)^\nu - \left(1 - \frac{1}{2^{k-1}}\right)^\nu.$$

This sums up to

$$G(z,u) = \sum_{\nu,k} \mathbf{P}(M=k) u^k \frac{z^\nu}{\nu!} = \sum_k u^k \left(e^{z(1-1/2^k)} - e^{z(1-1/2^{k-1})}\right). \tag{3}$$

Considering now the $m = 2^b$ buckets induces multinomials when distributing elements amongst buckets; therefore $n![z^n]G(z/m,u)^m$ is the probability generating function of $\sum_i M^{(i)}$.

The expressions for the first and second moment of Z are obtained from there by substituting respectively u by $2^{1/m}$ and $2^{2/m}$.

This gives the following lemma.

Lemma 1. *When there are n input items, the expected value and variance of the unnormalized estimator Z are*

$$\mathbf{E}(Z) = mn![z^n]G\left(\frac{z}{m}, 2^{1/m}\right)^m, \quad \textit{and} \tag{4}$$

$$\mathbf{Var}(Z) = m^2 n![z^n]G\left(\frac{z}{m}, 2^{2/m}\right)^m - \left(mn![z^n]G\left(\frac{z}{m}, 2^{1/m}\right)^m\right)^2. \tag{5}$$

3.2. **Poissonization.** If $f(z) = \sum_n f_n z^n/n!$ is the exponential generating function of the expectation of a parameter, the quantity $e^{-\lambda} f(\lambda) = \sum_n f_n e^{-\lambda}\lambda^n/n!$ gives the corresponding generating function under the Poisson model. Therefore the quantities

$$\mathcal{E}_n = mG\left(\frac{n}{m}, 2^{1/m}\right)^m (e^{-n/m})^m \quad \text{and} \quad \mathcal{V}_n = m^2 G\left(\frac{n}{m}, 2^{2/m}\right)^m e^{-n} - \mathcal{E}_n^2 \tag{6}$$

are respectively the mean and the variance of Z when the number of input items follows a Poisson law of rate $\lambda = n$.

We consider in the following the variable $\mathcal{E}_n$.

Using Equations 3 and 6, we can write

$$\mathcal{E}_n = mA(n)^m, \quad \text{where} \quad A(x) = \sum_i \frac{2^i}{m}\left(\phi(x/2^i) - \phi(x/2^{i-1})\right), \text{ and } \phi(x) = e^{-x/m}.$$

The Mellin transform $F^\star(s)$ (see [5, 8]) of a harmonic sum $F(x) = \sum \lambda_k f(\mu_k x)$ is

$$F^\star(s) = f^\star(s) \sum \frac{\lambda_k}{\mu_k^s};$$

this implies that

$$A^\star(s) = \phi^\star(s)(2^s - 1)\frac{2^{1/m}}{1 - 2^{1/m}2^s}.$$

The dominant singularity is at $s = -1/m$ and the corresponding residue is

$$a = m^{-1/m}\Gamma(-1/m)\frac{1 - 2^{1/m}}{\log 2}.$$

The Mellin transfer theorem gives the corresponding contribution $ax^{1/m}$ in the asymptotic expansion of $A(x)$ at infinity. The same techniques apply when considering $\mathcal{V}_n$. These results are summarized in the following lemma.

Lemma 2. *The Poisson mean $\mathcal{E}_n$ and variance $\mathcal{V}_n$ satisfy as $n \to \infty$:*

$$\mathcal{E}_n \sim \left[\left(\Gamma(-1/m)\frac{1-2^{1/m}}{\log 2}\right)^m + \eta_n\right] \times n, \tag{7}$$

$$\mathcal{V}_n \sim \left[\left(\Gamma(-2/m)\frac{1-2^{1/m}}{\log 2}\right)^m - \left(\Gamma(-1/m)\frac{1-2^{1/m}}{\log 2}\right)^{2m} + \kappa_n\right] \times n^2, \tag{8}$$

where $|\eta_n|$ and $|\kappa_n|$ (bounded by 10^{-6}) correspond to "negligible" singularities.

3.3. **Depoissonization.** The asymptotic forms of the first two moments of Z in the fixed-n model can be transferred back from the Poisson model by a method called "analytic depoissonization" by Jacquet and Szpankowski (See [7, 8]). The *values* of an exponential generating function at large arguments are closely related to the asymptotic form of its *coefficients* provided the generating function decays fast enough away from the positive real axis in the complex plane.

We have

$$G(z/m, 2^{1/m}) = e^{z/m} \sum_k 2^{k/m} e^{-z2^{-k}/m} \left(1 - e^{-z2^{-k}/m}\right).$$

Let S_θ be the cone

$$S_\theta = \{z : |\arg z| \le \theta\}, \quad \text{with } |\theta| < \pi/2.$$

There exists a θ such that

1. inside the cone S_θ there holds $e^{-z}G(z/m, 2^{1/m})^m = O(|z|)$, and
2. outside the cone S_θ there exists α such that $G(z/m, 2^{1/m})^m = O(e^{\alpha|z|})$.

This implies the following lemma (proof omitted).

Lemma 3. *The first two moments of the estimator Z are asymptotically equivalent under the Poisson and fixed-n model: $\mathbf{E}(Z) \sim \mathcal{E}_n$, $\mathbf{Var}(Z) \sim \mathcal{V}_n$.*

4. Improved Algorithm

An heuristic improvement consists in truncating the large non-meaningful values of the indicators M. When respectively 0%, 10%, 20% and 30% of the higher values are truncated, computations (for 32-bits words) give $\sigma \times \sqrt{S} = 2.8, 2.4, 2.2$, and 2.5.

Bibliography

[1] Alon (Noga), Matias (Yossi), and Szegedy (Mario). – The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, vol. 58, 1999, pp. 137–147.

[2] Durand (Marianne) and Flajolet (Philippe). – Loglog Counting of Large Cardinalities. – To be presented at ESA2003.

[3] Estan (Cristian) and Varghese (George). – New directions in traffic measurement and accounting. In *Proceedings of SIGCOMM 2002*. – ACM Press, 2002. (Also: UCSD technical report CS2002-0699, February, 2002; available electronically.).

[4] Flajolet (Philippe). – On adaptive sampling. *Computing*, vol. 34, 1990, pp. 391–400.

[5] Flajolet (Philippe), Gourdon (Xavier), and Dumas (Philippe). – Mellin transforms and asymptotics : Harmonic sums. *Theoretical Computer Science*, vol. 144, n° 1-2, 1995, pp. 3–58.

[6] Flajolet (Philippe) and Martin (G. Nigel). – Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, vol. 31, n° 2, 1985, pp. 182–209.

[7] Jacquet (Philippe) and Szpankowski (Wojciech). – Analytical depoissonization and its applications. *Theoretical Computer Science*, vol. 201, n° 1-2, 1998.

[8] Szpankowski (Wojciech). – *Average-Case Analysis of Algorithms on Sequences*. – John Wiley, New York, 2001.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 97–100.

Available online at the URL
http://algo.inria.fr/seminars/.

Quasi-Optimal Leader Election Algorithms in Radio Networks with Log-Logarithmic Awake Time Slots

Jean-François Marckert
Lama, University of Versailles (France)

April 28, 2003

Summary by Marianne Durand

Abstract

This summary presents two leader election protocols for radio networks, in both telephone and walkie-talkie model, where the number of radios is unknown. Those randomized protocols are shown to elect a leader in $O(\log n)$ expected time, and to be energy-efficient due to a small total awake time of the radio stations.

1. Introduction

A radio network consists in n radio stations with a graph of connection. The graph is assumed to be complete, so that all the stations are directly connected. All the stations are identical, and indistinguishable, and it is assumed that they are synchronized, and thus have a common notion of time. All stations can talk and listen. In a first model called *telephone model*, they can do both at the same time, and in a second called *walkie-talkie model* they can listen or talk but not both. When no station talk, there is a natural noise, when more than two stations talk, there is interference noise, whereas when only one station talks, all the other listening stations hear its message. The model of *collision detection* studied here is that a station that listens can distinguish between 1 and $\mathbb{N}/1$ for the number of emitting stations. The radio stations are supposed to run on batteries, and a station can be asleep in order to save them. In this case the station does not listen nor talk, and cannot be waken up by neighbours.

The question answered to in this summary, is *how to elect a leader?* That is how to elect a station such that all the n stations know that this station is the leader. This process should be fast and cheap in total awake time slots of the stations. If the total number of stations n is known, there is a trivial solution for the telephone model (that can be adapted for the walkie-talkie model) where each station broadcasts with probability $1/n$, and when only one station broadcasts, it is elected. This gives an election in $O(1)$ time. So from now on, n is supposed to be totally unknown.

The leader-election problem is fundamental in distributed systems, and has various applications in military communication and cellular phones. Algorithms that answer that problem can be found in [3], and Kushilevitz and Mansour in [1] gave a lower bound of $\Omega(\log n)$ for the average time of an election. The next section provides a class of efficients algorithms, that elect a leader in $O(\log n)$ expected time, with no station being awake for more than $O(\log\log n)$ time slots. Moreover, these algorithms feature a tuning parameter α that may be adjusted in order to reduce the election time, but with a larger total awake time for the stations, or at the opposite, reduce the energy consumption, but allowing a longer election. This summary is based on the paper [2].

2. Algorithms and Main Results

2.1. **Telephone.** This algorithm relies on the intuition that in general a station does not need to be awake when it does not broadcast, since each station that broadcasts listens, and can thus testify later of what happened. This suggests that each station must choose to be awake or asleep during a predetermined sequence of time slots, and then that all stations should wake up at a predetermined moment to hear the news about the election.

The general idea is that stations broadcast with probability $1/2, 1/4, \ldots$ until only one station broadcasts, but as this may never happen, we plan rounds of predetermined length, and at the end of each round, stations wake up to hear about the possible termination. The variable $\alpha > 1$ is a tuning parameter in the algorithm presented here.

Algorithm 'Tel'

round $\leftarrow 1$;
Repeat
  For k from 1 to $\lceil \alpha^{round} \rceil$ **do**
    Each station wakes up independently with probability $1/2^k$ to broadcast and to listen;
    If a unique station broadcasts **then** it becomes a *candidate* station;
  EndFor
  At the end of each round, all stations wake up and all candidates stations broadcast.
  If there is a unique candidate **then** it is *elected*
  round $\leftarrow$ *round* $+ 1$;
until a station is elected.

A station is a candidate if it is the only station to broadcast at a given time, and the algorithm ends if there is only one candidate in a round. The quality of this algorithm is based on the fast decrease inside a round of the broadcasting probability. Namely, if this probability 2^{-k} is much greater than $1/n$, for instance $k \leq \log_2 n - 3$, the probability to have more than two broadcasting stations is high. On the other side, if $k \geq \log_2 n + 3$, then the probability to have no broadcasting station is high. This shows that there is few time slots that may with good probability see a candidate. So that when *round* $\geq \log_\alpha \log_2 n$, the probability to have an election is significant (this probability is studied in the last section).

Theorem 1. *Let $q = 0.6308$, and let c be the function defined by*

$$c(\alpha, q) = \frac{q\alpha^3}{(\alpha - 1)(1 - \alpha(1 - q))}, \tag{1}$$

then, on average, Algorithm 'Tel' elects a leader in at most $c(\alpha, q) \log_2 n$ time slots, with no station being awake for more than $2 \log_\alpha \log_2 n(1 + o(1))$ mean time slots.

2.2. **Walkie-talkie.** For the walkie-talkie model, the algorithm has to face the problem that no candidate station can listen to its own message, therefore, the role of witness is introduced. A witness listens to the candidate station, and is later able to testify. Algorithm 'WT', that is close to Algorithm 'Tel', has a candidate iff at a time there is only one broadcasting station and only one witness. There is an election if there is only one candidate during the round.

Algorithm 'WT'

$round \leftarrow 1$;
Repeat
  For k from 1 to $\lceil \alpha^{round} \rceil$ **do**
    Each station wakes up independently with probability $1/2^k$;
    with probability 1/2 each awake station decides *either* to broadcast *or* to listen;
    a listening station that gets a message is a *witness*;
  EndFor
  At time $\lceil \alpha^{round} \rceil + 1$, each witness and each station that has broadcasted wakes up;
  Each witness broadcasts its received message;
  If there is a unique witness, the station that was witnessed is elected;
  At time $\lceil \alpha^{round} \rceil + 2$, all stations are listening;
    If a leader has been elected
      then the leader broadcasts and all the stations are aware of its status;
  $round \leftarrow round + 1$;
until a station is elected.

Theorem 2. *On the average, Algorithm 'WT' elects a leader in at most* $c(\alpha, q') \log_2 n$ *time slots, with no station being awake for more than* $2.5 \log_\alpha \log_2 n$ *mean time slots, with* $q' = 0.6176$.

3. Analysis

To analyze the properties of Algorithm 'Tel', like the average election time, we first study the probability to have an election at round j.

Proposition 1. *Let* p_j *denote the probability that a leader is elected at round* j*, and* $j(n)$ *be a sequence of integers such that* $n/2^{\alpha^{j(n)}} \to 0$*, then there exists* N *such that for all* $n > N$

$$p_{j(n)} \geq 0.3693.$$

The proof of this proposition is in two steps, first find a proper expression for p_j, second, find a lower bound.

There is an election at round j if there is only one candidate during this round, say at step k. So p_j is the sum on all possible k of the probability to have only one broadcasting station at step k, and no candidates during the other steps. The probability to have a candidate at step k is the probability that one station broadcasts (2^{-k}) times the probability that the others don't $((1-2^{-k})^{n-1})$ times the number of stations (n). So p_j is written as

$$\begin{aligned} p_j &= \sum_{k=1}^{\lceil \alpha^j \rceil} \frac{n}{2^k} \left(1 - \frac{1}{2^k}\right)^{n-1} \prod_{i \neq k} \left(1 - \frac{n}{2^i}\left(1 - \frac{1}{2^i}\right)^{n-1}\right) \\ &= \sum_{m=0}^{\infty} \sum_{k=1}^{\lceil \alpha^j \rceil} \left(\frac{n}{2^k}\left(1 - \frac{1}{2^k}\right)^{n-1}\right)^{m+1} \prod_{i=1}^{\lceil \alpha^j \rceil} \left(1 - \frac{n}{2^i}\left(1 - \frac{1}{2^i}\right)^{n-1}\right) \end{aligned}$$

Name s_j the product in the equation above, then by thin upper bounds, and numerical calculations, we get an upper bound for s_j. More precisely, using the notations of Proposition 1, we have that $s_{j(n)} \geq 0.1883$.

For the other terms in p_j, we first bound $(1-2^{-k})^{n-1} = e^{(n-1)\log(1-2^{-k})} \leq e^{-(n-1)/2^k}$, and then the Mellin transform gives us that

$$\sum_{k=1}^{\lceil \alpha^j \rceil} \left(\frac{n}{2^k}\right)^{m+1} e^{-(m+1)n/2^k} \sim \frac{(m+1)!}{(m+1)^{m+2}\log 2},$$

up to minor fluctuations. Then, summing on m and doing numerical calculations give the result of Proposition 1.

Proof of Theorem 1. The value $j^* = \lceil \log_\alpha \log_2 n \rceil$ plays a crucial role in the analysis. The probability of having an election before the round j^* is small, so we assume that no election happens before this round. Then we observe that $n/2^{\alpha^{j^*+1}} \to 0$ when $n \to \infty$, so that the results of Proposition 1 applies. As the probability to have an election after j^*+1 has a constant lower bound q, the average time of the election is smaller than q^{-1}. Finally we obtain that the average number of rounds n_1 needed to elect a leader has an average smaller than $j^* + q^{-1}$. The round number i last for a time $\lceil \alpha^i \rceil$, so if T_1 denotes the time of the election, we have

$$\mathbf{E}(T_1) \leq \mathbf{E}\sum_{i=1}^{n_1} \lceil \alpha^i \rceil \leq \sum_{k=1}^{\infty}\sum_{i=1}^{j^*+k} (1+\alpha^i) q(1-q)^{k-1} \leq c(\alpha, q)\log_2 n + O(\log\log n).$$

This proves the first part of Theorem 1. For the second part, simply observe that a station is awake less than once in a row on average, so as the number of rounds is bounded by $2\log_\alpha \log_2 n$, the total number of awakening time slots for a station is smaller than $2\log_\alpha \log_2 n$. □

Proof of Theorem 2. The proof of Theorem 2 follows the same main lines. It expresses the probability to have an election at round j, and gives a lower bound for this quantity. Then this bound is translated into an upper bound for the average election time. □

4. **Conclusion**

The two algorithms presented here offer a quasi-optimal (up to a constant factor) way to answer the leader election problem, in the telephone or walkie-talkie model. Moreover, they are energy-saving protocols.

The tuning parameter α allow to optimize the average time complexity or the awake time slots of the n stations.

Bibliography

[1] Kushilevitz (Eyal) and Mansour (Yishay). – An $\omega(d\log(n/d))$ lower bound for broadcast in radio networks. In *SIAM Journal on Computing*, vol. 27, pp. 702–712. – 1998.

[2] Lavault (Christian), Marckert (Jean-François), and Ravelomanana (Vlady). – Quasi optimal leader election algorithms in radio networks with log-logarithmic awake time slots. – 2003. To appear.

[3] Willard (Dan). – Log-logarithmic selection resolution protocols in a multiple access channel. *SIAM Journal of computing*, vol. 15, n° 2, 1986, pp. 468–477.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 101–104.

Available online at the URL
http://algo.inria.fr/seminars/.

Forty Years of 'Quicksort' and 'Quickselect': a Personal View

Conrado Martínez
Universitat Politècnica de Cataluny (Spain)

October 6, 2003

Summary by Marianne Durand

Abstract

The algorithms 'quicksort' [3] and 'quickselect' [2], invented by Hoare, are simple and elegant solutions to two basic problems: sorting and selection. They are widely studied, and we focus here on the average cost of these algorithms, depending on the choice of the sample. We also present a partial sorting algorithm named 'partial quicksort'.

1. 'Quicksort' and 'Quickselect'

The sorting algorithm 'quicksort' is based on the divide-and-conquer principle. The algorithm proceeds as follows. First a pivot is chosen, with a specified strategy. Then all the elements of the array but the pivot are compared to the pivot. The elements smaller than the pivot are stored before the pivot, and the elements larger after the pivot. These two sub-arrays are then recursively sorted.

We denote by C_n the average number of comparisons done to sort an array of size n, and $\pi_{n,k}$ the probability that the kth element is the chosen pivot in an array of size n. The recursive design of the algorithm is translated into a recurrence satisfied by the cost C_n:

$$C_n = n - 1 + t_n + \sum_{k=1}^{n} \pi_{n,k}(C_{k-1} + C_{n-k}). \tag{1}$$

The value t_n denote the cost, in terms of comparisons, of the choice of the pivot, that may depend on n.

The selection algorithm 'quickselect' is based on the same principles. To select the mth element out of an array of size n, first a pivot is chosen, then the array is partitioned into two sub-arrays, and the mth element is then recursively selected into the appropriate sub-array.

The average cost of selecting the mth element in an array of size n, in terms of comparisons, using this algorithm is denoted by $Q_{n,m}$, which satisfies the recurrence

$$Q_{n,m} = n - 1 + t_n + \sum_{k=1}^{m-1} \pi_{n,k} Q_{n-k,m-k} + \sum_{k=m+1}^{n} \pi_{n,k} Q_{k-1,m}. \tag{2}$$

In the particular case of the standard variant, where the pivot is chosen with a uniform probability ($\pi_{n,k} = 1/n$), those recurrences are solved, and lead to the theorem:

Theorem 1. *The average number of comparisons to sort n elements using the standard 'quicksort'* [3] *is*

$$C_n = 2(n+1)H_n - 4n \sim 2n \log n, \tag{3}$$

where H_k is the kth harmonic number. The average number of comparisons $Q_{n,m}$ to select the mth element out of n elements using the standard 'quickselect' [5] *is*

$$Q_{n,m} = 2((n+1)H_n - (n+3-m)H_{n+1-m} - (m+2)H_m + n + 3) = \Theta(n). \tag{4}$$

The maximum of the function $Q_{n,\alpha n}$ is located at $\alpha = 0.5$ and is worth $(2 + 2\log 2)n$.

2. **Different Sampling Strategies**

To improve the cost of these two algorithms, a first idea is to use a different algorithm for small subfiles (for example insertion sort), and a second idea is to use samples to select better pivots, and reduce the probability of uneven partitions which lead to quadratic worst case.

2.1. **Median of 3.** A simple strategy to select a pivot, due to Singleton [11], is *median of 3*. The pivot is chosen as the median of a sample of size 3, selected with uniform probability. The distribution of the pivot is now characterized by $\pi_{n,k} = \frac{(k-1)(n-k)}{\binom{n}{3}}$.

The average number of comparisons in this case is equal to ([10])

$$C_n = \frac{12}{7} n \log n + O(n),$$

which is roughly 14% less than standard 'quicksort'.

The median-of-3 strategy can also be applied to the algorithm 'quickselect'. Kirshenhofer, Martinez, and Prodinger studied this variant in [4]. By using bivariate generating functions and technical exact computation, they found that the average number of comparisons is

$$Q_{n,m} = 2n + \frac{72}{35} H_n - \frac{156}{35} H_m - \frac{156}{35} H_{n+1-m} + 3m - \frac{(m-1)(m-2)}{n} + O(1).$$

In the particular case where $m = \lceil n/2 \rceil$, the average number of comparisons is $\frac{11}{4}n + o(n)$, which is 18% less than in the standard case.

2.2. **Optimal sampling.** The median-of-3 strategy can be generalized to the median-of-$2t+1$ strategy for any integer t. Martínez and Roura in [6] consider the case where the size of the sample is $s = 2t + 1$, with $t = t(n)$ depending on n. Traditional techniques to solve recurrences cannot be used here. Their approach is to make an extensive use of the continuous master theorem of Roura [9]. This theorem states that if the sequence F_n satisfies the recursive equation

$$\left\{ \begin{array}{lcl} F_n & = & b_n \quad \text{if} n < N \\ F_n & = & t_n \sum_k w_{n,k} F_k \end{array} \right. , \tag{5}$$

with appropriate growth conditions on the $w_{n,k}$, then the asymptotic behavior of F_n is known. It depends on the growth of the toll function t_n and of the coefficients $w_{n,k}$, and is equivalent to $t_n \log^k n (\log\log n)^i$ or $n^\alpha \log^k n$, where all the coefficients involved are known. This theorem is the appropriate tool to deal with the recurrences satisfied by the cost of 'quicksort median of $2t+1$'.

The complexity considered in [6] is the *total cost*. The total cost is function of the number of comparisons and of the number of exchanges, and is defined by $\#comparisons + \zeta \#exchanges$, where ζ is usually considered to be around 4. We state the following theorem on the total cost of the algorithms 'quicksort' and 'quickselect'.

Theorem 2. *If we use samples of size s, with $s = o(n)$ and $s = \omega(1)$, then the average total cost of 'quicksort' is*

$$C_n = (1 + \zeta/4) n \log_2 n + o(n \log n). \tag{6}$$

The average total cost of 'quickselect', with the same sample strategy, to find an element of given random rank is

$$Q_n = 2(1 + \zeta/4)n + o(n). \tag{7}$$

The optimal sample size s^*, that minimizes the total cost of 'quicksort' and 'quickselect', satisfies $s^* = O(\sqrt{n})$, and depends on the number of comparisons done to select the pivot.

The conclusion is that if exchanges are expensive, we should use fixed-size samples and pick the median.

Many other strategies of pivot selection are available. For example 'median-3-3 quicksort', where the pivot is the median of three medians of three samples, each sample of size three [1]. This leads to all the strategies where the pivot is a median of other preselected medians, each issued of a selection strategy of the same type. The idea to keep in mind is that the gain obtained by a better pivot strategy should always be larger than the additional cost of the choice of the pivot.

2.3. **Adaptive sampling.** For the 'quicksort' algorithm, the best pivot is the median of the array. This is not obviously the best choice for 'quickselect', for example if m is small or close to n. The idea of Martínez, Panario, and Viola in [8] is to choose a pivot with relative rank in the sample close to $\alpha = m/n$.

3. **Partial Sort**

The partial sort problem is, given an array of size n, sort the m smallest elements. The algorithm 'quickselsort' answers this question. It selects the mth element by 'quickselect', and then applies 'quicksort' to the $m-1$ elements to its left. The cost of this algorithm is $\Theta(n + m \log m)$. Another way is the *heapsort-based partial sort*, that builds a heap and extracts m times the minimum. Its cost is also $\Theta(n + m \log m)$.

A solution, given by Martínez in [7], is 'partial quicksort'. This algorithm uses the principles of 'quicksort', and proceeds as follows. First find a pivot (with any strategy) and then recursively apply 'partial quicksort' to the sub-arrays concerned. More precisely, if the pivot is smaller than m, sort the left sub-array, and apply 'partial quicksort' to the right sub-array. If the pivot is greater than m, then apply 'partial quicksort' to the left sub-array.

The average number of comparisons $P_{n,m}$ needed to sort the m smallest elements in an array of size n satisfy the recurrence

$$P_{n,m} = n - 1 + t + \sum_{k=m+1}^{n} \pi_{n,k} P_{k-1,m} + \sum_{k=1}^{m} \pi_{n,k} \left(P_{k-1,k-1} + P_{n-k,m-k} \right).$$

In this recurrence, t is the number of comparisons done to choose the pivot, and k represents the position of the pivot, chosen with probability $\pi_{n,k}$. When k is greater than m, the algorithm sorts the m smallest elements in the left sub-array, that has size $k-1$. In the other case, when k is smaller than m, the algorithm sorts the entire left sub-array, and the $m-k$ smallest elements of the right sub-array, that has size $n-k$. We recognize that $P_{n,n}$ is the average cost of the algorithm 'quicksort'.

In the standard case, when $\pi_{n,k} = 1/n$, this recurrence is solved exactly, and we get that the average number of comparisons done by 'partial quicksort' is

$$P_{n,m} = 2n + 2(n+1)H_n - 2(n+3-m)H_{n+1-m} - 6m + 6. \tag{8}$$

'Partial quicksort' makes $2m - 4H_m + 2$ comparisons and $m/3 - 5H - m/6 + 1/2$ exchanges less than 'quickselsort'.

Bibliography

[1] Bentley (Jon L.) and McIlroy (Douglas). – Engineering a sort function. *Software—Practice and Experience*, vol. 23, n° 11, 1993, pp. 1249–1265.

[2] Hoare (Charles A. R.). – Find. *Communications of the ACM*, vol. 4, n° 7, 1961, pp. 321–322.

[3] Hoare (Charles A. R.). – Quicksort. *The Computer Journal*, vol. 5, n° 1, 1962, pp. 10–15.

[4] Kirschenhofer (P.), Prodinger (H.), and Martínez (C.). – Analysis of Hoare's FIND algorithm with median-of-three partition. *Random Structures & Algorithms*, vol. 10, n° 1–2, 1997, pp. 143–156.

[5] Knuth (Donald E.). – *The Art of Computer Programming*. – Addison-Wesley, 1998, seconde edition, vol. 3: Sorting and Searching.

[6] Martínez (C.) and Roura (S.). – Optimal sampling strategies in quicksort and quickselect. *SIAM Journal on Computing*, vol. 31, n° 3, 2001, pp. 683–705.

[7] Martinez (Conrado). – Partial quicksort. In *Proceedings of the First ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*. – 2004.

[8] Martínez (Conrado), Panario (Daniel), and Viola (Alfredo). – Adaptive sampling for quickselect. In *Proceedings of the 15thACM-SIAM Symp. on Discrete Algorithms (SODA)*. – 2004.

[9] Roura (Salvador). – An improved master theorem for divide-and-conquer recurrences. In *Automata, languages and programming (Bologna, 1997)*. pp. 449–459. – Springer, Berlin, 1997.

[10] Sedgewick (R.). – The analysis of quicksort programs. *Acta Informatica*, vol. 7, 1977, pp. 327–355.

[11] Singleton (Richard C.). – Algorithm 347: an efficient algorithm for sorting with minimal storage [m1]. *Communications of the ACM*, vol. 12, 1969, pp. 185 – 186.

Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 105–108.

Available online at the URL
http://algo.inria.fr/seminars/.

Overview of Sattolo's Algorithm

Mark C. Wilson
Department of Computer Science, University of Auckland (New Zealand)

June 21, 2004

Summary by Éric Fusy

Abstract

We give an overview of Sattolo's algorithm, which allows to perform random generation of a cyclic permutation of a fixed number of elements. In Section 1, we describe the algorithm and prove its correctness by using a recursive proof which parallelizes the recursive structure of the algorithm.

The recursive structure also allows to analyze two simple parameters associated to the algorithm. As we see in Section 2, simple recursive equations have been obtained by Prodinger and then studied by Mahmoud to obtain convergence results of the distribution of these parameters.

In Section 3, we present the method exposed by Mark C. Wilson in his talk to deal with the analysis of parameters associated to the algorithm. He uses a "grand" generating function $F(t, u, x)$ associated to each parameter and tries to obtain an explicit expression for this function. He only partially succeeds and finds an explicit expression for $\frac{\partial}{\partial x}\frac{F(t,u,x)}{x}$, from which Prodinger's and Mahmoud's results can be retrieved.

1. Sattolo's Algorithm

1.1. **Description.** In [4], Sattolo presents a very simple algorithm to uniformly sample a cyclic permutation σ of n elements.

The algorithm starts with the identity permutation $\sigma^{(0)} = \mathrm{Id}$. For each $i \in \{1, \ldots, n-1\}$, we denote by $\sigma^{(i)}$ the permutation obtained after the i first steps. Step i consists in choosing a random integer k_i in $\{1, \ldots, n-i\}$ and swapping the values of $\sigma^{(i-1)}$ at places k_i and $n-i+1$. In this way, we obtain a new permutation $\sigma^{(i)}$, which is equal to $\sigma^{(i-1)} \circ \tau_{k_i, n-i+1}$, where $\tau_{k_i,n-i+1}$ is the transposition exchanging k_i and $n-i+1$.

Finally, the algorithm returns the permutation $\sigma = \sigma^{(n-1)}$. An example of the execution of the algorithm is illustrated on Figure 1, where $n = 5$ and the sequence of chosen random integers is $3, 1, 2, 1$. The returned cyclic permutation on this example is $1 \to 5 \to 3 \to 2 \to 4 \to 1$.

Sattolo's algorithm is the adaptation for cyclic permutations of a very well known algorithm [1] to sample a permutation of n elements at random. The only difference is that Sattolo's algorithm chooses the random integer k_i in $\{1, \ldots, n-i\}$ whereas the algorithm of [1] chooses k_i randomly in $\{1, \ldots, n-i+1\}$ at step i.

1.2. **Correctness.** The fact that the algorithm returns a uniformly distributed random cyclic permutation follows from the unicity and existence of the decomposition of a cyclic permutation σ as a product $\tau_{k_1,n} \circ \cdots \circ \tau_{k_i,n-i+1} \circ \cdots \circ \tau_{k_{n-1},2}$, where $k_i \in \{1, \ldots, n-i\}$ for $1 \le i \le n-1$.

$$\begin{array}{ccccc} 1 & 2 & \underline{3} & 4 & \mathbf{5} \\ \underline{1} & 2 & 5 & \mathbf{4} & 3 \\ 4 & \underline{2} & \mathbf{5} & 1 & 3 \\ \underline{4} & \mathbf{5} & 2 & 1 & 3 \\ 5 & 4 & 2 & 1 & 3 \end{array}$$

FIGURE 1. The execution of Sattolo's algorithm for $n = 5$, and when the sequence of chosen random swapping places is $3, 1, 2, 1$.

This property can be established recursively by associating to σ the number $q(\sigma) = \sigma(n)$. As σ is cyclic, $q(\sigma) \in \{1, \dots n-1\}$. In addition $\tau_{q(\sigma),n} \circ \sigma$ fixes n and is cyclic on $\{1, \dots, n-1\}$: indeed, with the cyclic notation, if $\sigma = (n, q(\sigma), r_1, \dots, r_{n-3})$, then $\tau_{q(\sigma),n} \circ \sigma = (n)(q(\sigma), r_1, \dots, r_{n-3})$.

Notation. We denote the permutation $\tau_{q(\sigma),n} \circ \sigma$ by $\sigma_\downarrow$.

2. Analysis of the Algorithm: Probabilistic Approaches

In the literature, two parameters are analyzed: the number of times a value k is moved is denoted by $M_{n,k}$ and the total distance covered by a value k is denoted by $D_{n,k}$. For example, on Figure 1, the values of $M_{n,k}$ are $1, 1, 1, 2, 3$ and the values of $D_{n,k}$ are $3, 1, 2, 4, 4$ for $k = 1, 2, 3, 4, 5$.

2.1. Prodinger's approach.

In [3], Prodinger introduces the probabilistic generating function $\phi_{n,k}(u) = \sum_l P(M_{n,k} = l)u^l$ associated to the parameter $M_{n,k}$ and the probabilistic generating function $\xi_{n,k}(u) = \sum_l P(D_{n,k} = l)u^l$ associated to the parameter $D_{n,k}$.

Using the recursive structure of the algorithm, he obtains a recursive system of two equations for $\phi_{n,k}(u)$:

$$\left\{ \begin{array}{rcl} \phi_{n,k}(u) & = & \frac{n-k}{n-1}u + \frac{k-1}{n-1}\phi_{k,k}(u) \quad 1 \le k < n \\ \phi_{n,n}(u) & = & \frac{u}{n-1}\sum_{k=1}^{n-1}\phi_{n-1,k}(u) \quad n \ge 2, \quad \phi_{1,1}(u) = 1 \end{array} \right. \tag{1}$$

We note $E_{n,k} = \mathbf{E}(M_{n,k})$. Observing that $E_{n,k} = \phi'_{n,k}(1)$ and derivating Equation-system 1 at $u = 1$, we find the following recursive system:

$$\left\{ \begin{array}{rcl} E_{n,k} & = & \frac{n-k}{n-1} + \frac{k-1}{n-1}E_{k,k} \\ E_{n,n} & = & 1 + \frac{1}{n-1}\sum_{k=1}^{n-1}E_{n-1,k} \end{array} \right. \tag{2}$$

From this system, it is easy to deduce an explicit expression for $\mathbf{E}(M_{n,k})$:

$$\mathbf{E}(M_{n,k}) = \frac{n+2k-5}{n-1} \quad k \ge 2, \ \mathbf{E}(M_{n,1}) = 1 \ n \ge 2, \ \mathbf{E}(M_{1,1}) = 0 \tag{3}$$

Similarly, we can find an explicit expression for $\mathbf{Var}(M_{n,k})$:

$$\mathbf{Var}(M_{n,k}) = \frac{2(k-2)(3n+1-2k)}{(n-1)^2} - \frac{4H_{k-2}}{n-1} \quad k \ge 2, \ \mathbf{Var}(M_{n,1}) = 0 \tag{4}$$

For the parameter $D_{n,k}$, the recursive structure of the algorithm yields:

$$\left\{ \begin{array}{rcl} \xi_{n,k}(u) & = & \frac{u^{n-k}}{n-1} + \frac{n-2}{n-1}\xi_{n-1,k}(u), \quad 1 \leq k < n \\ \xi_{n,n}(u) & = & \frac{1}{n-1}\sum_{k=1}^{n-1} \xi_{n-1,k}(u)u^{n-k} \quad n \geq 2,\ \xi_{1,1}(u) = 1 \end{array} \right. \tag{5}$$

From these equations, we can also obtain an exact expression for the mean and variance of the variable $D_{n,k}$.

2.2. **Mahmoud's refinements.** In [2], Mahmoud considers the "randomized" variable M_{n,K_n} where K_n is a random element uniformly chosen in $\{1, \ldots, n\}$. Writing $\psi_n(u) = \mathbf{E}(u^{M_{n,K_n}})$ for its probabilistic generating function, he obtains from Equation-system 1 the simple recursive equation $\psi_n(u) = \frac{n-2+u}{n}\psi_{n-1}(u) + \frac{u}{n}$. Hence $\psi_n(u) - \frac{u}{2-u} = \frac{n-2+u}{n}\left(\psi_{n-1}(u) - \frac{u}{2-u}\right)$. As a consequence, he obtains

$$\psi_n(u) = \frac{u}{2-u}\left(1 - \frac{2\Gamma(n-u+1)}{u\Gamma(u-1)\Gamma(n+1)}\right) \tag{6}$$

Thus, for $0 \leq u < 2$ and according to Stirling formula, $\psi_n(u) \to_{n\to\infty} \frac{1}{2}\frac{u}{1-u/2}$, which is the probabilistic generating function of a geometric random variable Geo(1/2). As a consequence, M_{n,K_n} converges in distribution to Geo(1/2), a result which can be intuitively predicted from the recursive structure of the algorithm.

Then Mahmoud "derandomizes" M_{n,K_n}, using the equation $\phi_{n,k}(u) = \frac{n-k}{n-1}u + \frac{k-1}{n-1}u\psi_k(u)$. He finds an explicit limit $\phi_\alpha(u)$ for $\phi_{n,k}(u)$ when $\frac{k}{n} \to_{n\to\infty} \alpha$, such that $\phi_\alpha(u)$ is the probabilistic generating function of a random variable $X_\alpha = B + (1-B)(1 + \mathrm{Geo}(\frac{1}{2}))$ where B has law $\mathrm{Ber}(\alpha)$. Hence, when $\frac{k}{n} \to_{n\to\infty} \alpha$, $M_{n,k}$ converges in distribution to a mixture of the constant 1 and of the random variable $1 + \mathrm{Geo}(\frac{1}{2})$, where the random variable mixing the two variables is a Bernoulli law $\mathrm{Ber}(\alpha)$ with rate α. The mean and variance of this random variable agree with the limit of the exact expressions of Prodinger for $\mathbf{E}(M_{n,k})$ and $\mathbf{Var}(M_{n,k})$ when $\frac{k}{n} \to_{n\to\infty} \alpha$.

Similarly, Mahmoud randomizes the problem for $\xi_{n,k}$. He considers the random variable D_{n,K_n}, where K_n is an integer uniformly distributed in $\{1, \ldots, n\}$. A scaling is necessary to obtain a convergence result. We have to consider the variable $\widetilde{D}_{n,K_n} = \frac{1}{n}(D_{n,K_n} - K_n)$. Writing $\widetilde{\eta}(u)$ for the probabilistic generating function of $\widetilde{D}_{n,K_n}$, Mahmoud finds that $\widetilde{\eta}(e^t) \to_{n\to\infty} \int_0^1 \frac{e^{\theta t} - e^{-\theta t}}{2\theta t} d\theta$. Hence, $\widetilde{D}_{n,K_n}$ converges in distribution to a product of two independant uniform $U(0,1)$ and $U(-1,1)$ random variables, a less intuitive result than for the case of M_{n,K_n}.

3. The Method of Mark Wilson

3.1. **Introduction.** Mark Wilson wants to generalize the approach of Prodinger and Mahmoud to analyze Sattolo's algorithm. He prefers to associate a "grand" combinatorial generating function rather than probabilistic generating functions, although both approaches can easily be linked as we will see. His method is presented in [5].

Noting $\mathcal{C} = \cup_n \mathcal{C}_n$ the set of cyclic permutations, $n(\sigma)$ the number of elements permuted by a cyclic permutation, and $\chi(\sigma, p)$ a parameter associated to σ such as $M_{n(\sigma),p}$ or $D_{n(\sigma),p}$, he introduces the "grand" generating function

$$F(u,t,x) = \sum_{\sigma\in\mathcal{C}, p\in[n(\sigma)]} u^{\chi(\sigma,p)} t^p \frac{x^{n(\sigma)}}{|\mathcal{C}_{n(\sigma)}|}.$$

Observe that

$$\begin{aligned} F(u,t,x) &= \sum_{n\geq 1} \frac{x^n}{(n-1)!} \sum_{1\leq p\leq n} t^p \sum_{\sigma\in\mathcal{C}_n} u^{\chi(\sigma,p)} \\ &= \sum_{n\geq 1} x^n \sum_{1\leq p\leq n} t^p \phi_{n,p}^{(\chi)}(u) \end{aligned}$$

where $\phi_{n,p}^{(\chi)}(u)$ is the probabilistic generating function associated to $\chi(\sigma,p)$. This establishes a link between Prodinger's probabilistic notations and these notations.

3.2. **Originality and advantages.** The originality of Mark Wilson's method is that it tries to establish an exact expression for the "grand" generating function $F(t,u,x)$. From such an expression, the results of Prodinger and Mahmoud could be easily retrieved. In addition, extended results could be obtained such as the analysis of the algorithm when the cyclic permutations are not uniformly distributed on $\mathcal{C}_n$ or even when the size of the random cyclic permutation is a random variable, such as a geometric variable for example. As we will see, the method does not completely succeed.

3.3. **Description of the method on an example.** We treat here the case where $\chi(\sigma,p) = M_{n(\sigma),p}$. First, exact recursive formulae for $\chi(\sigma,p)$ are obtained. These formulae are groundly equivalent to the recursive probabilistic formulae of Prodinger:

$$\chi(\sigma,p) = \begin{cases} \chi(\sigma_\downarrow,p) & \text{if } p\neq n(\sigma), p\neq q(\sigma); \\ 1+\chi(\sigma_\downarrow,p) & \text{if } p= n(\sigma), p\neq q(\sigma); \\ 1 & \text{if } p\neq n(\sigma), p= q(\sigma); \\ 0 & \text{if } p= n(\sigma), p= q(\sigma). \end{cases} \tag{7}$$

We denote by $\{\mathcal{I}_1,\mathcal{I}_2,\mathcal{I}_3,\mathcal{I}_4\}$ the partition of $\mathcal{I} = \{(\sigma,p)/\sigma\in\mathcal{C}, 1\leq p\leq n(\sigma)\}$ induced by Equation 7. For each $i\in\{1..4\}$, we denote by $\Sigma_i(t,x,u)$ the associated generating function. Equation 7 can easily be translated in 4 equations involving the Σ_i.

This system of 4 equations can be solved and yields the following expression for $\frac{\partial}{\partial x}\frac{F(u,t,x)}{x}$:

$$(1-x)\frac{\partial}{\partial x}\frac{F(u,t,x)}{x} = ut^2\frac{u}{2-u}\frac{1}{(1-tx)^2} + \frac{2(1-u)}{2-u}(1-tx)^{-u} + \frac{ut}{1-t}\left(\frac{1}{1-x}-\frac{t}{1-tx}\right) \tag{8}$$

from which Prodinger's and Mahmoud's results can easily be retrieved. Unfortunately, this expression can not be easily integrated to give an explicit expression for $F(u,t,x)$.

A similar treatment can be carried out to deal with $D_{n,k}$.

Bibliography

[1] Knuth (Donald E.). – *Sorting and Searching*. – Addison-Wesley, 1973, *The Art of Computer Programming*, vol. 3.

[2] Mahmoud (Hosam M.). – Mixed distributions in sattolo's algorithm for cyclic permutations via randomization and derandomization. *Journal of Applied Probability*, vol. 40, 2003, pp. 790–796.

[3] Prodinger (Helmut). – On the analysis of an algorithm to generate a random cyclic permutation. *Ars Combinatoria*, vol. 65, 2002, pp. 75–78.

[4] Sattolo (Sandra). – An algorithm to generate a random cyclic permutation. *Information Processing Letters*, vol. 22, 1986, pp. 315–317.

[5] Wilson (Mark C.). – Probability generating functions in sattolo's algorithm. – `http://www.cs.auckland.ac.nz/~mcw/Research/Mypapers/papers.html`.

Contents

Part I. Enumerative Combinatorics

Part II. Analytic Combinatorics and Asymptotics

Part III. Analysis of Algorithms and Protocols

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS
Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
(France)
http://www.inria.fr
ISSN 0249-6399