Algorithms for Analytic Combinatorics – PI4 Program 2018

Stephen Melczer

One of the draws of combinatorics is its ability to take inspiration from, motivate, and even push forward diverse areas of mathematics and computer science. In particular, much current research focuses on the use of analytic techniques to address questions of computability and complexity in enumerative combinatorics. The universality of many analytic statements often allows for very general enumerative results which are ripe for automation. This program will study these techniques and their wide range of applications across several disciplines, with a focus on implementing algorithms which will be of use to current and future researchers.

Background

The basic tool of enumerative combinatorics is the generating function. Given a sequence $(f_n) = f_0, f_1, \dots,$ the *generating function* of (f_n) is the formal power series

$$F(z) = \sum_{n\geq 0} f_n z^n.$$

There are well established methodologies going from a combinatorial description of some family of objects to a specification of the generating function counting the number of objects of a given size. Furthermore, when $F(z)$ represents the Taylor series of an analytic function at the origin there is a strong link between analytic properties of $F(z)$ and asymptotics of the sequence (f_n). Below we outline some examples of the theory, more details will be given in beginning lectures and can be found in Flajolet and Sedgewick [1].

Example 1. The Cauchy root test for series convergence implies that if a minimum modulus singularity of $F(z)$ has modulus ρ then the *exponential growth* of (f_n) is

$$\limsup_{n\to\infty} |f_n|^{1/n} = 1/\rho.$$

Example 2. Let $B(z)$ denote the generating function for the number b_n of binary trees on n nodes. The recursive definition of a rooted tree as a leaf or a node with two ordered binary tree children implies

$$B(z) = z + zB(z)^2,$$

and the quadratic formula gives

$$B(z) = \frac{1-\sqrt{1-4z^2}}{2z}$$

(the other quadratic root is irrelevant). As the smallest (non-removable) singularity is at $1/4$, b_n grows like 4^n times something subexponential. In fact, the generalized Newton binomial theorem implies $b_n = \frac{1}{n+1}\binom{2n}{n}$, the Catalan number of order n.

Example 3. An *ascending permutation* of size $2k+1$ is a rearrangement of the numbers $1, 2, \dots, 2k+1$ such that from left to right the elements alternate increasing and decreasing (so the first is smaller than the second, the second is larger than the third, etc.) Let f_n denote the number of ascending permutations of size n (zero when n is even) and $F(z)$ denote its *exponential generating function* $F(z) = \sum_{n\geq 0} \frac{f_n}{n!} z^n$. Using methods for manipulating generating functions, it can be shown that $F(z)$ satisfies the differential equation

$$\frac{dF}{dz}(z) = 1 + F(z)^2, \qquad F(0) = 0.$$

Solving this differential equation implies that $F(z) = \tan(z)$. (!)[1] Furthermore, since the smallest modulus singularities of $\tan(z)$ are simple poles at $\pm\pi/2$, the methods of analytic combinatorics immediately imply that $f_n \sim n! \cdot 2(2/n)^{n+1}$ (for n odd).

Example 4. Euclid's GCD algorithm over the finite field F_p works by taking two polynomials u_0 and u_1 with $\deg(u_1) < \deg(u_0)$ and performing successive divisions until a remainder of zero is reached. Let $\mathcal{F}_n$ denote the set of pairs of polynomials (u_0, u_1) in F_p with u_0 monic and $\deg(u_1) < \deg(u_0) = n$. Let $f_{n,k}$ denote the number of elements of $\mathcal{F}_n$ on which the Euclidean algorithm takes k steps to terminate. A short but clever argument implies that the *multivariate generating function* $F(z, u)$ satisfies

$$F(z, u) = \sum_{n,k\geq 0} f_{n,k} z^n u^k = \frac{1}{1 - pz - up(p-1)z}.$$

Because the singularities of $F(z, u)$ satisfy certain easily checked conditions, the distribution of the number of steps of Euclid's algorithm on a random pair of polynomials of size n (i.e., in $\mathcal{F}_n$) limits to a normal distribution with mean $(1 - 1/p)n + O(1)$ and variance $O(n)$ as $n \to \infty$.

These examples just begin to scratch the surface of the strong, striking results possible using the theory of analytic combinatorics. Below we list potential projects that students will work on; other projects may be possible depending on the backgrounds and interests of participants. Questions can be directed to *smelczer@sas.upenn.edu*.

References

[1] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009, p. 810.

Project 1: Algorithms for Analytic Combinatorics in Several Variables

Given a multivariate rational function with power series expansion

$$F(\mathbf{z}) = \sum_{\mathbf{i}\in\mathbb{N}^d} f_{\mathbf{i}} \mathbf{z}^{\mathbf{i}} = \sum_{\mathbf{i}\in\mathbb{N}^d} f_{i_1,\dots,i_d} z_1^{i_1} \cdots z_d^{i_d},$$

the *(complete) diagonal* of $F(\mathbf{z})$ is the univariate function $(\Delta F)(z) = \sum_{n\geq 0} f_{n,\dots,n} z^n$, whose coefficients are known as the *diagonal sequence* of $F(\mathbf{z})$. Diagonal sequences arise frequently in combinatorics (lattice path enumeration, statistics on trees, irrational tilings), probability theory (random walk models), number theory (binomial sums such as Apéry's sequence, used in his proof of the irrationality of $\zeta(3)$), operations research (queuing theory), and physics (statistical mechanics) – see Melczer [1] or Pemantle and Wilson [3] for many examples. Given a rational function $F(\mathbf{z})$, one often wants to determine asymptotics of its diagonal sequence or classify its algebraic properties (for instance, determining when the diagonal is transcendental). The aim of this project will be to produce software that takes as input a rational function $F(\mathbf{z})$ satisfying various assumptions and outputs dominant asymptotics of its diagonal sequence.

[1] This was first noticed by Désiré André in 1881. See the first chapter of Flajolet and Sedgewick [1] for details.

The first phase of this project will be to implement recent symbolic-numeric algorithms described by Melczer and Salvy [2, 1] in Sage (or an equivalent language). These algorithms rely on numerical computations which must be certified in order to be rigorous, potentially through the use of interval arithmetic.

The results of Melczer and Salvy require assumptions which hold *generically* among rational functions of fixed numerator and denominator degree, but which often do not hold in practice. The second phase of this project will allow group members to relax such assumptions and develop new theoretical algorithms. Depending on how the project proceeds, and participants backgrounds, results on the effective stratification of algebraic varieties, real algebraic geometry, and other areas of mathematics, can be incorporated.

References

[1] Stephen Melczer. "Analytic Combinatorics in Several Variables: Effective Asymptotics and Lattice Path Enumeration". Thesis (PhD). PhD thesis. University of Waterloo and ENS Lyon, 2017, 259 pages. URL: https://arxiv.org/abs/1709.05051.

[2] Stephen Melczer and Bruno Salvy. "Symbolic-Numeric Tools for Analytic Combinatorics in Several Variables". In: *ISSAC 2016—Proceedings of the 41st International Symposium on Symbolic and Algebraic Computation* (2016).

[3] R. Pemantle and M. C. Wilson. "Twenty combinatorial examples of asymptotics derived from multivariate generating functions". In: *SIAM Review* 50 (2008), pp. 199–272.

Project 2: Algorithms for Algebraic Generating Functions

Algebraic generating functions are ubiquitous in many areas of combinatorics, theoretical computer science, and bioinformatics (for example, the generating function of any unambiguously generated context-free language is algebraic). The goal of this project is to develop algorithms which take a sequence with algebraic generating function $F(z)$, specified by its minimal polynomial $P(F, z) = 0$ and some initial terms, and output dominant asymptotics of the sequence. Following the theory of analytic combinatorics, one must determine the singularities of the generating function $F(z)$, together with information about the types of singularities (for example, are the singularities poles, algebraic branch points, etc.).

It is classical that any singularity of $F(z)$ must be a root of either the *discriminant* of P or the leading coefficient of P as a polynomial in $\mathbb{Z}[z]$, however these algebraic sets will contain additional elements corresponding to the other solutions of $P(F, z) = 0$ (see Figure 1). In the first phase of this project, students will implement homotopy-based algorithms for algebraic coefficient asymptotics using previous theoretical results [3, 6].

Every algebraic function satisfies a linear differential equation with polynomial coefficients, with good bounds on the order and coefficient degrees which arise [2]. Group members will compare the practicality of their algorithms with current software for determining asymptotics of solutions to such differential equations [7].

Finally, every algebraic function is the diagonal of a bivariate rational function [4] and, conversely, every bivariate rational diagonal is an algebraic function (with good bounds and algorithms known [1]). In later stages of this project, group members can implement a recent algorithm for asymptotics of bivariate rational diagonals [5] (different from the approach detailed in Project 1) and compare the complexity of these approaches to asymptotics of algebraic generating functions.

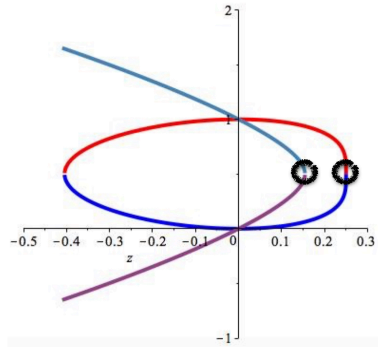

Figure 1: The generating function $F(z)$ of *bi-coloured supertrees* satisfies $F^4-2F^3+(1+2z)F^2-2zF+4z^3=0$. Which branch corresponds to the generating function, and which singular point determines asymptotics?

References

[1] Alin Bostan, Louis Dumont, and Bruno Salvy. "Algebraic diagonals and walks: algorithms, bounds, complexity". In: *J. Symbolic Comput.* 83 (2017), pp. 68–92. ISSN: 0747-7171. URL: https://doi.org/10.1016/j.jsc.2016.11.006.

[2] Alin Bostan et al. "Differential equations for algebraic functions". In: *ISSAC 2007*. ACM, New York, 2007, pp. 25–32. DOI: 10.1145/1277548.1277553. URL: http://dx.doi.org/10.1145/1277548.1277553.

[3] Cyril Chabaud. "Séries génératrices algébriques : asymptotique et applications combinatoires". Type : Thèse de Doctorat – Soutenue le : 2002-11-22 – Dirigée par : Soria, Michèle. PhD thesis. LIP6, 2002.

[4] J. Denef and L. Lipshitz. "Algebraic power series and diagonals". In: *J. Number Theory* 26.1 (1987), pp. 46–67. ISSN: 0022-314X. DOI: 10.1016/0022-314X(87)90095-3. URL: http://dx.doi.org/10.1016/0022-314X(87)90095-3.

[5] T. DeVries, J. van der Hoeven, and R. Pemantle. "Automatic asymptotics for coefficients of smooth, bivariate rational functions". In: *Online J. Anal. Comb.* 6 (2011), 24 pages.

[6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009, p. 810.

[7] Marc Mezzarobba. *Rigorous Multiple-Precision Evaluation of D-Finite Functions in SageMath*. Tech. rep. 1607.01967. Extended abstract of a talk at the 5th International Congress on Mathematical Software. arXiv, 2016.

Project 3: Algorithms for Rational Function Decomposition

When dealing with multivariate rational functions there is no canonical generalization of the classical partial fraction decomposition for univariate rational functions. One such generalization, due to Leǐnartas, is the following [1, p. III.17] [4].

Proposition 5. *Let K be a field with algebraic closure $\overline{K}$ and $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z})$ with $G, H \in K[\mathbf{z}]$ coprime. Let $H = H_1^{r_1} \cdots H_m^{r_m}$ be the irreducible factorization of H in $K[\mathbf{z}]$ and $\mathcal{V}_i := \{\mathbf{z} \in \overline{K}^d : H_i(\mathbf{z}) = 0\}$. Then $F(\mathbf{z})$ can be written in the form*

$$F(\mathbf{z}) = \sum_{\mathcal{A}} \frac{F_A(\mathbf{z})}{\prod_{i \in \mathcal{A}} G_i^{s_i}},$$

where the s_i are positive integers (possibly greater than the r_i), each $F_A(\mathbf{z})$ is a (possibly zero) polynomial in $K[\mathbf{z}]$, and the sum is taken over all subsets $\mathcal{A} \subset \{1, \ldots, m\}$ such that $\cap_{i \in \mathcal{A}} \mathcal{V}_i \neq \varnothing$ and $\{G_i : i \in \mathcal{A}\}$ is algebraically independent over $K[\mathbf{z}]$.

The aim of this project is to come up with efficient algorithms to compute such a decomposition, and compare them to a naive approach (already implemented in Sage [3]). There are currently enumeration problems coming from representation theory, among other areas, where such an algorithm is the limiting factor in computing desired asymptotics.

As a starting point, participants will attempt to create an efficient algorithm for the linear case.

Proposition 6. *Let $F(\mathbf{z}) = G(\mathbf{z})/H(\mathbf{z})$ with $G, H \in K[\mathbf{z}]$ co-prime and $H = l_1^{r_1} \cdots l_m^{r_m}$ for linear functions l_j. Then $F(\mathbf{z})$ can be written in the form*

$$F(\mathbf{z}) = \sum_{\mathcal{A}} \frac{F_A(\mathbf{z})}{\prod_{i \in \mathcal{A}} l_i^{s_i}},$$

where the s_i are positive integers (possibly greater than the r_i), each $F_A(\mathbf{z})$ is a (possibly zero) polynomial in $K[\mathbf{z}]$, and the sum is taken over all subsets $\mathcal{A} \subset \{1, \ldots, m\}$ such that $\{l_i : i \in \mathcal{A}\}$ is linearly independent.

The linear case finds application to problems from queuing theory [2].

Note that such decompositions are not unique, and participants will also search for relationships between the different decompositions. On a related topic, a given generating function $F(z)$ can be the diagonal of an infinite number of multivariate rational functions[2]. Although there is an effective procedure for determining when two multivariate rational functions have the same diagonal, it is less clear how to characterize classes of functions with the same diagonal. If up for a challenge, participants could investigate characterizations of rational functions with equivalent diagonals and try to determine canonical members of equivalence classes which are suitable for an asymptotic analysis.

References

[1] I. A. Aĭzenberg and A. P. Yuzhakov. *Integral representations and residues in multidimensional complex analysis.* Vol. 58. Translations of Mathematical Monographs. Translated from the Russian by H. H. McFaden, Translation edited by Lev J. Leifman. American Mathematical Society, Providence, RI, 1983, pp. x+283. ISBN: 0-8218-4511-X.

[2] Andrea Bertozzi and James McKenna. "Multidimensional residues, generating functions, and their application to queueing networks". In: *SIAM Rev.* 35.2 (1993), pp. 239–268. ISSN: 0036-1445. DOI: 10.1137/1035045. URL: http://dx.doi.org/10.1137/1035045.

[3] A. Raichev. *amgf Documentation – Release 0.8.* https://github.com/araichev/amgf. 2012.

[4] A. Raichev. "Leinartas's partial fraction decomposition". In: *arxiv:'1206.4740'* (2012).

[2] There are at least 5 different diagonal representations of Apéry's sequence for the irrationality of $\zeta(3)$ in the literature!