SIAM REVIEW
Vol. 50, No. 2, pp. 199–272

Twenty Combinatorial Examples of Asymptotics Derived from Multivariate Generating Functions*

Robin Pemantle†
Mark C. Wilson‡

Abstract. Let $\{a_{\mathbf{r}} : \mathbf{r} \in \mathbb{N}^d\}$ be a d-dimensional array of numbers for which the generating function $F(\mathbf{z}) := \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is meromorphic in a neighborhood of the origin. For example, F may be a rational multivariate generating function. We discuss recent results that allow the effective computation of asymptotic expansions for the coefficients of F. Our purpose is to illustrate the use of these techniques on a variety of problems of combinatorial interest. The survey begins by summarizing previous work on the asymptotics of univariate and multivariate generating functions. Next we describe the Morse-theoretic underpinnings of some new asymptotic techniques. We then quote and summarize these results in such a way that only elementary analyses are needed to check hypotheses and carry out computations. The remainder of the survey focuses on combinatorial applications, such as enumeration of words with forbidden substrings, edges and cycles in graphs, polyominoes, and descents in permutations. After the individual examples, we discuss three broad classes of examples, namely, functions derived via the transfer matrix method, those derived via the kernel method, and those derived via the method of Lagrange inversion. These methods have the property that generating functions derived from them are amenable to our asymptotic analyses, and we describe further machinery that facilitates computations for these classes of examples.

Key words. stratified Morse theory, residue, enumeration, smooth point, multiple point, singularity, kernel method, Lagrange inversion, transfer matrix

AMS subject classifications. Primary, 05A15, 05A16; Secondary, 32A05

DOI. 10.1137/050643866

Contents.

1 Introduction **201**
1.1 Background: The Univariate Case 202
1.2 Multivariate Asymptotics 202
1.3 New Multivariate Methods 204

2 A Brief Review of Univariate Methods **207**
2.1 Saddle Point Methods 207
2.2 Circle Methods 208

*Received by the editors October 31, 2005; accepted for publication (in revised form) January 29, 2007; published electronically May 5, 2008.
http://www.siam.org/journals/sirev/50-2/64386.html

†Department of Mathematics, University of Pennsylvania, 209 S. 33rd Street, Philadelphia, PA 19104 (pemantle@math.upenn.edu). The research of this author was supported in part by National Science Foundation grant DMS 0103635.

‡Department of Computer Science, University of Auckland, Private Bag 92019, Auckland, New Zealand (mcw@cs.auckland.ac.nz).

2.3 Transfer Theorems 209

3 New Multivariate Results **212**
3.1 Topological Representation 213
3.2 Geometric Criteria 215
3.3 Asymptotic Formulae for Minimal Smooth Points 221
3.4 Multiple Point Asymptotics 222
3.5 Distributional Limits 224

4 Detailed Examples **226**
4.1 Binomial Coefficients 227
4.2 Delannoy Numbers 228
4.3 Powers, Quasi-powers, and Generalized Riordan Arrays 230
4.4 Maximum Number of Distinct Subsequences 232
4.5 Paths, Hills, and Fine Numbers 235
4.6 Horizontally Convex Polyominoes 239
4.7 Symmetric Eulerian Numbers 241
4.8 Smirnov Words 243
4.9 Alignments of Sequences 244
4.10 Probability That There Is an Edge in an Induced Subgraph 245
4.11 Integer Solutions to Linear Equations 248
4.12 Queuing Theory 249

5 Transfer Matrices **251**
5.1 Restricted Switching 252
5.2 Connector Matrices 253
5.3 Forbidden Substring Example 254

6 Lagrange Inversion **256**
6.1 Bivariate Asymptotics 257
6.2 Trees in a Simple Variety 258
6.3 Cores of Planar Graphs 259

7 The Kernel Method **260**
7.1 A Random Walk Problem 260
7.2 Explanation of the Kernel Method 262
7.3 Dyck, Motzkin, Schröder, and Generalized Dyck Paths 263
7.4 Pebble Configurations 266

8 Discussion of Other Methods **268**
8.1 GF-Sequence Methods 268
8.2 The Diagonal Method 269

List of Generating Functions by Subsection

4.1	$\frac{1}{1-x-y}$
4.2	$\frac{1}{1-x-y-xy}$
4.4	$\frac{1}{1-x-xy(1-x)^d}$
4.5	$\frac{2}{1+2x+\sqrt{1-4x}-2xy}$
4.6	$\frac{xy(1-x)^3}{(1-x)^4-xy(1-x-x^2+x^3+x^2y)}$
4.7	$\frac{e^x-e^y}{xe^y-ye^x}$
4.8	$\frac{1}{1-\sum_{j=1}^d \frac{z_j}{1+z_j}}$
4.9	$\frac{1}{2-\prod_{j=1}^k (1+z_j)}$
4.10	$\frac{1}{1-z(1-x^2y^2)} \frac{1}{1-x(1+y)}$
4.12	$\frac{\exp(x+y)}{(1-\rho_{11}x-\rho_{21}y)(1-\rho_{12}x-\rho_{22}y)}$
5.1	$\frac{P(u,v,z)}{1+z^3v+uz^3+uz^2+uz^2v+z^2v-uz-2z-zv-uz^4v}$
5.2	$\left[1-(x_1+\cdots+x_d)-\operatorname{trace}\left((\mathbf{I}+\mathbf{V})^{-1}\mathbf{LJ}\right)\right]^{-1}$
5.3	$\frac{1+x^2y^3+x^2y^4+x^3y^4-x^3y^6}{1-x-y+x^2y^3-x^3y^3-x^4y^4-x^3y^6+x^4y^6}$
6.3	$\frac{3xz(1-z)(3-z)}{(1-3y(1+z)^2)(27-xz(3-z)^2)}$
7.1	$\frac{1}{1+\sqrt{1-x-y}}$
7.3	$\frac{2}{1+\sqrt{1-4x^2}-2xy}$
7.3	$\frac{2}{1+\sqrt{1-2x-3x^2}-x-2xy}$
7.3	$\frac{2}{1+\sqrt{1-6x^2+x^4}-x^2-2xy}$
7.4	$\frac{\eta(x,y)}{x-(x+y)^2}$
8.2	$\frac{1+xy+x^2y^2}{1-x-y+xy-x^2y^2}$

1. Introduction.

The purpose of this paper is to review recent developments in the asymptotics of multivariate generating functions, and to give an exposition of them that is accessible and centered around applications. The introductory section lays out notions of generating functions and their asymptotics, and delimits the scope of this survey.

We employ the standard asymptotic notation: "$f = O(g)$" is taken to mean $\limsup_{z\to z_0} |f(z)/g(z)| < \infty$, where the limit z_0 is specified but does not appear in the notation, "$f = o(g)$" means $f/g \to 0$, and "$f \sim g$" means $f/g \to 1$, again accompanied by a specification of which variable is bound and the limit to which it is taken. When we say "$a_n = O(g(n))$" we always mean as $n \to \infty$; for multivariate arrays and functions, statements such as "$a_{\mathbf{r}} = O(g(\mathbf{r}))$" must be accompanied by a specification of how $\mathbf{r}$ is taken to infinity. We typically use $\mathbf{r}$ for $(r_1, \ldots, r_d) \in \mathbb{Z}^d$

and $\mathbf{z}$ for $(z_1, \ldots, z_d) \in \mathbb{C}^d$, but sometimes it is clearer to use (r, s, t) for (r_1, r_2, r_3) or (x, y, z) for (z_1, z_2, z_3). The norm $|\mathbf{r}|$ of a vector $\mathbf{r}$ denotes the L^1 norm $\sum_j |r_j|$.

1.1. Background: The Univariate Case. Let $\{a_n : n \geq 0\}$ be a sequence of complex numbers and let $f(z) := \sum_n a_n z^n$ be the associated generating function. For example, if a_n is the nth Fibonacci number, then $f(z) = 1/(1 - z - z^2)$. Generating functions are among the most powerful tools in combinatorial enumeration. In the introductory section of his graduate text [66], Richard Stanley deems a generating function $f(z)$ to be "the most useful but most difficult to understand method for evaluating" a_n, compared to a recurrence, an asymptotic formula, and a complicated explicit formula.

There are two steps involved in using generating functions to evaluate a sequence: first, one must determine f from the combinatorial description of $\{a_n\}$, and second, one must be able to extract information about a_n from f. The first step is partly science and partly an art form. Certain recurrences for $\{a_n\}$ translate neatly into functional equations for f, but there are numerous twists and variations. Linear recurrences with constant coefficients, such as in the Fibonacci example, always lead to rational generating functions, but often there is no way to tell in advance whether f will have a sufficiently nice form to be useful. A good portion of many texts on enumeration is devoted to the battery of techniques available for producing the generating function f.

The second step, namely, estimation of a_n once f is known, is reasonably well understood and somewhat mechanized. Starting with Cauchy's integral formula

$$a_n = \frac{1}{2\pi i} \int z^{-n-1} f(z)\, dz\,, \tag{1.1}$$

one may apply complex analytic methods to obtain good estimates for a_n. As a prelude to the multivariate results which are the main subject of this survey, we will give a quick primer on the univariate case (section 2 below), covering several well-known ways to turn (1.1) into an estimate for a_n when n is large.

1.2. Multivariate Asymptotics. In 1974, when Bender published the review article [6], the extraction of asymptotics from multivariate generating functions was largely absent from the literature. Bender's concluding section urges research in this area:

> Practically nothing is known about asymptotics for recursions in two variables even when a generating function is available. Techniques for obtaining asymptotics from bivariate generating functions would be quite useful.

By the time of Odlyzko's 1995 survey [55], a single vein of research had appeared, initiated by Bender and carried further by Gao, Richmond, and others. Let the number of variables be denoted by d so that $\mathbf{z}$ denotes the d-tuple $(z_1, \ldots, z_d)$. We use the multi-index notation for products: $\mathbf{z}^\mathbf{r} := \prod_{j=1}^d z_j^{r_j}$. Suppose we have a multivariate array $\{a_\mathbf{r} : \mathbf{r} \in \mathbb{N}^d\}$ for which the generating function

$$F(\mathbf{z}) = \sum_\mathbf{r} a_\mathbf{r} \mathbf{z}^\mathbf{r}$$

is assumed to be known in some reasonable form. We are interested in the asymptotic behavior of $a_\mathbf{r}$. For concreteness, one may keep in mind the following examples, which are discussed in detail in sections 4.1 and 4.2.

EXAMPLE 1.1 (binomial coefficients). *Let* $a_{rs} = \binom{r+s}{r,s}$. *The array* $\{a_{rs}\}$ *is Pascal's triangle, oriented so that the rays of ones emanate from the origin along the positive* r *and* s *axes. From the recursion* $a_{rs} = a_{r-1,s} + a_{r,s-1}$, *holding whenever* $(r,s) \neq (0,0)$, *one easily obtains*

$$F(x,y) := \sum_{r,s \geq 0} a_{rs} x^r y^s = \frac{1}{1-x-y}\,.$$

EXAMPLE 1.2 (Delannoy numbers). *Let* a_{rs} *count the number of paths from the origin to the lattice point* (r,s) *made of three kinds of steps:* 1 *unit east,* 1 *unit north, and* $\sqrt{2}$ *units northeast. These are called* Delannoy numbers. *The recursion* $a_{r,s} = a_{r-1,s} + a_{r,s-1} + a_{r-1,s-1}$ *leads immediately to*

$$F(x,y) := \sum_{r,s \geq 0} a_{rs} x^r y^s = \frac{1}{1-x-y-xy}\,.$$

In both of these examples, F is a rational function. In one variable, obtaining asymptotics for the coefficients of a rational function is a quick exercise (see (2.1) below) but in more than one variable this is far from true. For example, whereas the binomial coefficients are exact expressions easily approximated via Stirling's formula, the Delannoy numbers do not have such a simple expansion (the expression $a_{rs} = \sum_i \binom{r}{i}\binom{s}{i} 2^i$ is readily derived from the generating function, but no simpler one is forthcoming).

In [67, Example 6.3.8], it is shown how to get more information about the *central* Delannoy numbers a_{nn} by finding the generating function $\sum_n a_{nn} z^n$. Uniform estimation of the general term a_{rs} is not possible by this diagonal method (see section 8.2), though the problem succumbs easily to multivariate methods (see (1.4) below). The class of multivariate generating functions addressed in this survey is larger than the rational functions; the exact hypothesis is that the function is meromorphic in a certain domain, which is spelled out in the remark following Theorem 3.16.

The first paper to concentrate on extracting asymptotics from multivariate generating functions was [5], already published at the time of Bender's survey, but the seminal paper is [7]. In this paper, Bender and Richmond assume that F has a singularity of the form $A/(z_d - g(\mathbf{x}))^q$ near the graph of a smooth function g, for some real exponent q, where $\mathbf{x}$ denotes $(z_1, \ldots, z_{d-1})$. They show, under appropriate further hypotheses on F, that the probability measure μ_n obtained by renormalizing $\{a_{\mathbf{r}} : r_d = n\}$ to sum to 1 converges to a multivariate normal when appropriately rescaled. Their method, which we call the *GF-sequence method*, is to break the d-dimensional array $\{a_{\mathbf{r}}\}$ into a sequence of $(d-1)$-dimensional slices and consider the sequence of $(d-1)$-variate generating functions

$$f_n(\mathbf{x}) = \sum_{\mathbf{r}: r_d = n} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}\,.$$

They show that, asymptotically as $n \to \infty$,

$$f_n(\mathbf{x}) \sim C_n g(\mathbf{x}) h(\mathbf{x})^n \tag{1.2}$$

uniformly over $\mathbf{x}$ in a certain ball, and that sequences of generating functions obeying (1.2) satisfy a central limit theorem and a local central limit theorem.

We will review these results in more detail in section 8, but one crucial feature is that they always produce Gaussian (central limit) behavior. The applicability of the entire method is therefore limited to the single, though important, case where the coefficients $a_{\mathbf{r}}$ are nonnegative and possess a Gaussian limit. The work of [7] has been greatly expanded upon, but always in a similar framework. For example, it has been extended to matrix recursions [9] and the applicability has been extended from algebraic to algebraico-logarithmic singularities of the form $F \sim (z_d - g(\mathbf{x}))^q \log^\alpha(1/(z_d - g(\mathbf{x})))$ [24]. The difficult step is always deducing asymptotics from the hypotheses (1.2). Thus some papers in this stream refer to such an assumption in their titles, and the term "quasi-power" has been coined for such a sequence $\{f_n\}$.

The theory has also been advanced via its use in applications. The forthcoming textbook by Flajolet and Sedgewick [21] devotes a chapter of nearly 100 pages to multivariate asymptotics, in which many of the basic results on quasi-powers are reviewed and extended. The reader is referred to this substantial body of work for an extensive collection of applications that can be handled by reducing multivariate problems to univariate contour integrals. The limit theorems in [21, Chapter IX] (outside the graph enumeration example in section 11) are all Gaussian. The large deviation results of Hwang [41, 42] are not restricted to the Gaussian case, but give asymptotics on a cruder scale.

1.3. New Multivariate Methods. Odlyzko's survey of asymptotic enumeration methods [55], which is meant to be somewhat encyclopedic, devotes fewer than 6 of its 160 pages to multivariate asymptotics. Odlyzko describes why he believes multivariate coefficient estimation to be difficult. First, the singularities are no longer isolated, but form $(d-1)$-dimensional hypersurfaces. Thus, he points out, "Even rational multivariate functions are not easy to deal with." Second, the multivariate analogue of the one-dimensional residue theorem is the considerably more difficult theory of Leray [45]. This theory was later fleshed out by Aĭzenberg and Yuzhakov, who spend a few pages [1, section 23] on generating functions and combinatorial sums. Further progress in using multivariate residues to evaluate coefficients of generating functions was made by Bertozzi and McKenna [10], though at the time of Odlyzko's survey none of the papers based on multivariate residues such as [47, 10] had resulted in any kind of systematic application of these methods to enumeration.

The topic of the present review article is a recent vein of research begun in [57] and continued in [58, 4, 50] and in several manuscripts in progress. The idea, seen already to some degree in [10], is to use complex methods that are genuinely multivariate to evaluate coefficients of multivariate generating functions via the multivariate Cauchy formula. By avoiding symmetry-breaking decompositions such as $F(\mathbf{z}) = \sum f_n(z_1, \ldots, z_{d-1}) z_d^n$, one hopes that the methods will be more universally applicable and the formulae more canonical. In particular, the results of Bender et al. and the results of Bertozzi and McKenna are seen to be two instances of a more general result that estimates the Cauchy integral via topological reductions of the cycle of integration. These topological reductions, while not fully automatic, are algorithmically decidable in large classes of cases and are the subject of section 3. An ultimate goal, stated in [57, 58], is to develop computer software to automate all of the computation.

Aside from providing a summary and explication of this line of research, the present survey is meant to serve several other purposes. First, results from [57, 58, 4] are presented in streamlined forms, stated so as to avoid the scaffolding one needs to prove them. This is to make the results more comprehensible. Second, by focusing on combinatorial applications, we hope to create a sort of user's manual, one that contains worked examples akin to those a potential user will have in mind. Many of the applications are to abstract combinatorial structures, but we also include direct applications to computational biology and to formal languages and automata theory. Finally, we present a number of results that "preprocess" the basic, general theorems, providing useful computational reductions of the hypotheses or conclusions in specific cases of interest. We now give the notation for generating functions and their asymptotics that will be used throughout the paper, and then briefly describe a prototypical asymptotic theorem.

Throughout this survey, we let

$$F(\mathbf{z}) = \frac{G(\mathbf{z})}{H(\mathbf{z})} = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}} \tag{1.3}$$

be a generating function in d variables, where G and H are analytic and $H(\mathbf{0}) \neq 0$. Recall that in the bivariate case ($d = 2$) we write

$$F(x, y) = \frac{G(x, y)}{H(x, y)} = \sum_{r,s=0}^{\infty} a_{rs} x^r y^s \,.$$

The representation of F as a quotient of analytic functions is required to hold only in a certain domain, described in the remark following Theorem 3.16, though in the majority of the examples F is meromorphic on all of $\mathbb{C}^d$. We will assume throughout that H vanishes somewhere, since the methods in this paper do not give nontrivial results for entire functions.

We are concerned with asymptotics when $|\mathbf{r}| \to \infty$ with $\hat{\mathbf{r}} := \mathbf{r}/|\mathbf{r}|$ remaining in some specified set, bounded away from the coordinate planes. Thus, for example, when $d = 2$, the ratio s/r will remain in a compact subset of $(0, \infty)$. It is possible via our methods to address the other case, where $r = o(s)$ or $s = o(r)$ (see, for example, [50]), but our main purpose in this paper is to give examples that require, among all the methods and results cited above, only those from [57] together with the simplest methods from [58, 4].

To illustrate the sort of basic results we quote from [57, 58], we give the following combination of Corollary 3.18 and Theorem 3.20 from section 3.

THEOREM 1.3. *Let F be as in* (1.3) *and suppose $a_{\mathbf{r}} \geq 0$.*

(i) *For each $\mathbf{r}$ in the positive orthant, there is a unique $\mathbf{z}(\mathbf{r})$ in the positive orthant satisfying the equations $r_d z_j \partial H/\partial z_j = r_j z_d \partial H/\partial z_d$ $(1 \leq j \leq d-1)$ from* (3.5) *below and lying on the boundary of the domain of convergence of F; the quantity $\mathbf{z}(\mathbf{r})$ depends on $\mathbf{r}$ only through the direction $\mathbf{r}/|\mathbf{r}|$.*

(ii) *With $\mathbf{z}(\mathbf{r})$ defined in this way, if $G(\mathbf{z}(\mathbf{r})) \neq 0$,*

$$a_{\mathbf{r}} \sim (2\pi)^{-(d-1)/2} \mathcal{H}^{-1/2} \frac{G(\mathbf{z}(\mathbf{r}))}{-z_d \partial H/\partial z_d(\mathbf{z}(\mathbf{r}))} (r_d)^{-(d-1)/2} \mathbf{z}(\mathbf{r})^{-\mathbf{r}}$$

uniformly over compact cones of $\mathbf{r}$ for which $\mathbf{z}(\mathbf{r})$ is a smooth point of $\{H = 0\}$ uniquely solving (3.5) *on the boundary of the domain of convergence of F, and for which $\mathcal{H}$ is nonzero, where $\mathcal{H}$ is the determinant of the Hessian matrix of the function parametrizing the hypersurface $\{H = 0\}$ in logarithmic coordinates.*

Going back to Example 1.1, let us see what this says about binomial coefficients. It will turn out (see section 4.1) that $\mathbf{z}(\mathbf{r}) = (\frac{r}{r+s}, \frac{s}{r+s})$. Evaluating the Hessian and the partial derivatives then leads to (4.1) below:

$$a_{rs} \sim \left(\frac{r+s}{r}\right)^r \left(\frac{r+s}{s}\right)^s \sqrt{\frac{r+s}{2\pi rs}}$$

as $r, s \to \infty$ at comparable rates. This agrees, of course, with Stirling's formula. When we try this with the Delannoy numbers from Example 1.2, we find that

$$\mathbf{z}(r, s) = \left(\frac{\sqrt{r^2+s^2}-s}{r}, \frac{\sqrt{r^2+s^2}-r}{s}\right)$$

and that

(1.4)
$$a_{rs} \sim \left(\frac{\sqrt{r^2+s^2}-s}{r}\right)^{-r} \left(\frac{\sqrt{r^2+s^2}-r}{s}\right)^{-s} \sqrt{\frac{1}{2\pi}} \sqrt{\frac{rs}{(r+s-\sqrt{r^2+s^2})^2\sqrt{r^2+s^2}}}$$

uniformly as $r, s \to \infty$ with r/s and s/r remaining bounded. Setting $r = s = n$ gives the estimate

$$a_{nn} \sim (\sqrt{2}-1)^{-2n} \sqrt{\frac{1}{2\pi n} \frac{2^{-1/4}}{2-\sqrt{2}}}$$

for the central Delannoy numbers. This estimate can also be computed, with more effort, by using the methods of section 2.3 applied to the diagonal generating function, which can itself be derived using the *diagonal method* as in [67, Example 6.3.8] (we discuss the diagonal method in section 8.2).

In the sections to follow, we work through a number of applications of this and other newer theorems to problems in combinatorial enumeration. While doing so, we develop companion results that simplify the computations in certain classes of examples. A prototypical companion result is the following result for implicitly defined generating functions, which arise commonly in recursions on trees.

If $f(z) = z\phi(f(z))$, then the nth coefficient of f is given asymptotically by

$$n^{-3/2} \frac{y_0 \phi'(y_0)^n}{\sqrt{2\pi\phi''(y_0)/\phi(y_0)}}, \tag{1.5}$$

where y_0 is the unique positive solution to $y\phi'(y) = \phi(y)$.

In [21] the computation is reduced via Lagrange inversion to the determination of the y^{n-1} coefficient of $\phi^n(y)$, which is then carried out via complex integration methods. By viewing this coefficient instead as the (n, n) coefficient of

$$F(x, y) := \frac{y}{1 - x\phi(y)},$$

one sees that the complex integration step is already done and that (1.5) is immediate upon identifying that $\mathbf{z}(n, n) = (1/\phi(y_0), y_0)$; see Proposition 6.1, where the hypotheses for (1.5) are stated more completely.

The organization of the remainder of this paper is as follows. The next section gives the promised quick primer in univariate methods. Section 3 is the most theoretical, outlining results from various sources to give a brief but nearly complete explanation as to how one goes from (1.3), via a multivariable Cauchy formula, to asymptotic formulae for $a_{\mathbf{r}}$. We then quote the precise theorems we will use from [57] and [58]. While we attempt here to provide short and accessible statements of results, readers interested in knowing only enough to handle a particular application may wish to skip this section, find the closest match among the examples in sections 4–7, and then refer back to the necessary parts of section 3. Section 4 works in detail through a number of diverse examples. The subsequent three sections discuss collections of examples arising from three combinatorial methods: transfer matrices, Lagrange inversion, and the kernel method, respectively; we present applications to enumerating various kinds of words, paths, trees, and graphs. In section 8 we discuss open questions and extensions of the material presented here, and compare our results with those of other authors.

2. A Brief Review of Univariate Methods.

Since our main subject is the asymptotics of coefficients of multivariate generating functions, we will give only a quick overview of the univariate case touching on three widely used methods, namely, saddle point analysis, Darboux' circle method, and branch point contours. In each case we will give a brief summary and an application or two, as well as some pointers to the literature. For readers who wish to understand univariate asymptotic methods in greater detail, we recommend beginning with sections 10–12 of Odlyzko's survey paper [55].

We begin by disposing of a trivial case. If $f(z)$ is a rational function, written as $P(z)/Q(z)$ for polynomials P and Q, then a partial fraction decomposition exists:

$$f(z) = \sum_{\alpha,k} \frac{P_{\alpha,k}(z)}{(1-z/\alpha)^k}, \tag{2.1}$$

where α ranges over roots of Q and the integer k is between 1 and the multiplicity of the root α. For $j \geq i$, the z^j coefficient of $\frac{z^i}{(1-z/\alpha)^k}$ is $\alpha^{i-j}\binom{j-i+k}{k}$, which leads to a representation

$$a_n = \sum_{\alpha} p_\alpha(n)\alpha^{-n}$$

for easily computed polynomials p_α.

We turn now to analyses via contour integrals. To see what underlies all of these methods, we begin with a preliminary estimate. Suppose that f has radius of convergence R. Taking the contour of integration in (1.1) to be a circle of radius $R - \epsilon$ gives $a_n = O(R-\epsilon)^{-n}$ for any $0 < \epsilon < R$. If f is continuous on the closed disk of radius R, then $a_n = O(R^{-n})$. All the methods we discuss refine these basic estimates by pushing the contour out far enough that the resulting upper bound becomes asymptotically sharp.

2.1. Saddle Point Methods.

We begin with a worked example, then comment on the method in general.

EXAMPLE 2.1 (set partitions). *Let a_n be the number of (unordered) partitions of $\{1,\ldots,n\}$ into ordered sets. This has exponential generating function* [62, p. 194]

$$f(z) := \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n = e^{z/(1-z)}\,.$$

To evaluate the Cauchy integral, we attempt to move the contour so it passes through a point where the integrand is not rapidly oscillating. Specifically, we let $I_n =$

$\log(z^{-n-1}f(z))$ *denote the logarithm of the integrand and we find where* I_n' *vanishes. It is not hard to see that*

$$I_n' = \frac{-n-1}{z} + \frac{1}{(1-z)^2}$$

vanishes at a value $1-\beta_n$*, where* $\beta_n = n^{-1/2} + O(n^{-1})$*. Expand the contour to a circle passing through* $1-\beta_n$ *or, equivalently and more cleanly, a piece of the line* $1-\beta_n+ix$ *near* $x=0$*. Replacing the integrand* $\exp(I_n(z))\,dz$ *by its two-term Taylor approximation, one obtains*

$$\begin{aligned}\frac{a_n}{n!} &= \frac{1}{2\pi i}\int_{-\infty}^{\infty}\exp(I_n(1-\beta_n+it))(i\,dt)\\ &\sim \frac{\exp(I_n(1-\beta_n))}{2\pi}\int_{-\infty}^{\infty}\exp\left(\frac{1}{2}I_n''(1-\beta_n)(it)^2\right)\,dt\\ &= \frac{\exp(I_n(1-\beta_n))}{2\pi}\sqrt{\frac{1}{2\pi I_n''(1-\beta_n)}}\,.\end{aligned}$$

The approximation may be justified by routine estimates. Plugging the value of $1-\beta_n$ *into* I *and* I_n'' *shows the estimate to be asymptotically*

$$(1+o(1))\sqrt{\frac{1}{4\pi e n^{3/2}}}\exp(2\sqrt{n})\,.$$

For this method to work either f must be entire or the saddle point (where I_n' vanishes) must be in the interior of the domain of convergence of f. These conditions are frequently satisfied and the method is widely applicable. We do not know when saddle point approximation was first used in the context of generating functions. Hayman's seminal paper from 1956 [38] defines a broad class of functions, called *admissible functions*, for which the saddle point method can be shown to work and the Gaussian approximation mechanized. A variation of the generating function $\exp(z/(1-z))$ is analyzed in Hayman's Theorem XIII.

2.2. Circle Methods. Suppose f has a positive, finite radius of convergence, R. If there is no saddle point of $z^{-n-1}f(z)$ in the open disk of radius R, one might try to push the contour of integration near or onto the circle of radius R. Darboux' method is essentially this, with the refinement that if f extends to the circle of radius R and is k times continuously differentiable there, then

$$\int z^{-n}f(z)\,dz = O\left(n^{-k}R^{-n}\right)\,. \tag{2.2}$$

This follows from integration by parts.

The following very old theorem may be found, among other places, in [39, Theorem 11.10b: "Theorem of Darboux"].

Theorem 2.2 (Darboux). *Let* $f(z) = \sum_n a_n z^n = (r-z)^\alpha L(z)$*, where* $r>0$*,* α *is not a positive integer, and* L *is analytic in a disk of radius greater than* r*. Then*

$$a_n \sim r^{\alpha-n}n^{-\alpha-1}\frac{L(r)}{\Gamma(-\alpha)}\,.$$

EXAMPLE 2.3 (2-regular graphs). *Let b_n be the number of* 2*-regular graphs. The exponential generating function* [71, equation (3.9.1)] *is given by*

$$f(z) := \sum_n \frac{b_n}{n!} z^n = \frac{e^{-z/2-z^2/4}}{\sqrt{1-z}} .$$

Letting $\alpha = -1/2, r = 1$, and $L(z) = e^{-z/2-z^2/4}$ gives

$$a_n := \frac{b_n}{n!} \sim \frac{e^{-3/4}}{\sqrt{\pi n}} .$$

To prove the above version of Darboux' theorem, write $L(z)$ as m terms of a Taylor series about r plus a remainder. This expresses $f(z)$ as

$$\sum_{k=0}^{m-1} c_k (r-z)^{\alpha+k} + O\left((r-z)^{\alpha+m}\right) .$$

The known asymptotics for coefficients of $(r-z)^\nu$ together with the estimate (2.2) is good enough to yield $m-1$ terms in the conclusion provided one takes $m \geq 1 + (\mathrm{Re}\,\{-\alpha\})^+$, where x^+ denotes the maximum of x and 0. ◻

There are a great many variations on this, depending on how badly f behaves near the circle of radius r. One of the most classical fruits of the circle method is Hardy and Ramanujan's estimate [36] of the partition numbers.

EXAMPLE 2.4. *Let p_n denote the number of partitions of n, that is, representations of n as a sum of positive integers with the summands written in descending order. The number of identities involving these numbers and their relatives is staggering (see, e.g.,* [2] *for an elementary survey). Euler observed that the generating function may be written as*

$$\lambda(z) := \sum_{n=0}^{\infty} p_n z^n = \prod_{k=0}^{\infty} \frac{1}{1-z^k} .$$

From this, Hardy and Ramanujan obtained

$$p_n \sim \frac{\exp(\pi\sqrt{2n/3})}{4\sqrt{3}n}$$

as $n \to \infty$.

2.3. Transfer Theorems. A closer look at the proof of Darboux' theorem shows that one can do better. Analyticity of $f/(r-z)^\alpha$ beyond the disk of radius r is used only to provide a series for f in decreasing powers of $(r-z)$. On the other hand, taking the contour to be a circle of radius r loses one power of n in the estimate, though this doesn't matter when the expansion of f has sufficiently many terms.

In light of these observations, one sees that choosing a custom-built contour can simultaneously weaken the analyticity assumption while strengthening the estimate. The choice of contour will depend on the nature of the singularity of f on the boundary of its domain of convergence, but in many instances a good choice is known. Both a nonintegral power of $(r-z)$ and a logarithm or iterated logarithm of $1/(r-z)$ are branch singularities and require similar contours. In the simplest case, the contour, pictured in Figure 1, consists of an arc of a circle of radius $1/n$ around r, an arc

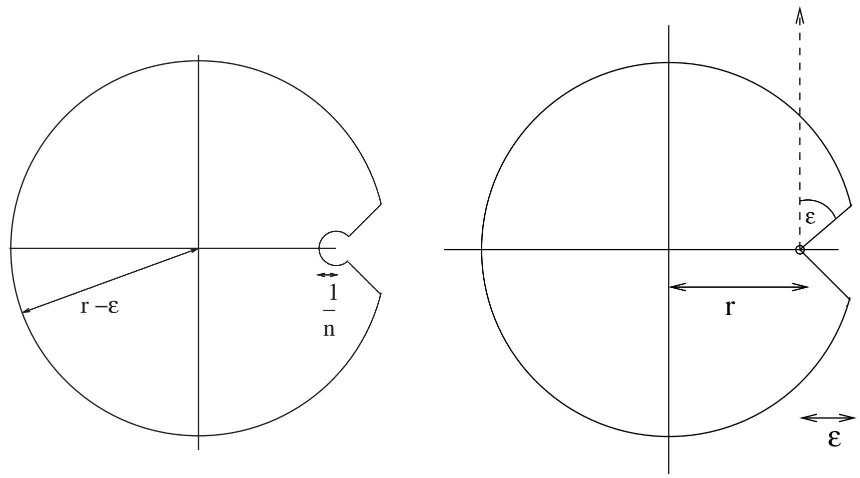

Fig. 1 *Contour for an algebraic singularity and the corresponding region of analyticity.*

of a circle of radius $r - \epsilon$ centered at the origin, and two line segments connecting corresponding endpoints of the two circular arcs.

Among the many refinements of Darboux' theorem via such contours are the *transfer theorems* of Flajolet and Odlyzko. Their paper [20] is masterfully written and worth taking a couple of pages here to summarize. They begin by isolating the gist of (2.2), namely, that $a_n = O(n^{-\lfloor \alpha \rfloor})$ if $f(z) = O(1-z)^\alpha$. Already this allows for f not to be in any nice class of functions, as long as it is bounded by a class we understand. Next, they use the above contour to improve this to a sharp estimate,

$$f(z) = O(1-z)^\alpha \quad \Rightarrow \quad a_n = O(n^{-\alpha-1})\,. \tag{2.3}$$

Let alg-log denote the class of functions that are a product of a power of $r - z$, a power of $\log(1/(r-z))$, and a power of $\log\log(1/(r-z))$. Flajolet and Odlyzko estimate a_n for any function in alg-log. Their last crucial observation is that the implication in (2.3) gives an explicit constant for the bound on a_n based on the constant in the bound $f(z) = O(1-z)^\alpha$, which is sufficiently constructive to give

$$f(z) = o\left((r-z)^\alpha\right) \quad \Rightarrow \quad a_n = o(n^{-\alpha-1})\,.$$

The surprisingly strong consequence of this is that any f with an asymptotic expansion by functions $\{g_k\}$ in alg-log has coefficients asymptotically given by summing the asymptotics of the coefficients of the functions g_k.

The main result of [20] begins by stating what analyticity assumption is necessary in order to use the contour in Figure 1. Given a positive real R and $\epsilon \in (0, \pi/2)$, the so-called *Camembert-shaped region*,

$$\{z : |z| < R + \epsilon, z \neq R, |\arg(z - R)| \geq \pi/2 - \epsilon\}\,,$$

shown on the right of Figure 1, is defined so that it includes the contour in Figure 1.

THEOREM 2.5. *Let $f(z) = \sum a_n z^n$ be singular at R but analytic in a Camembert-shaped region. For $g(z) = \sum b_n z^n$ in the class* alg-log, *asymptotic relations between f and g as $z \to R$ imply estimates on the coefficients as follows:*

(i) $f(z) = O(g(z)) \Rightarrow a_n = O(b_n)$.
(ii) $f(z) = o(g(z)) \Rightarrow a_n = o(b_n)$.
(iii) $f(z) \sim g(z) \Rightarrow a_n \sim b_n$.

In particular, when $f(z) \sim C(r-z)^\alpha$, this result subsumes Theorem 2.2.

We have indicated the arguments for deriving the second and third parts of the theorem from the first. We now sketch, in the simplest case where $g(z) = (1-z)^\alpha$, the upper estimate on a_n, which appears as Corollary 2 of [20]. The integral over the large circular arc is small because z^{-n-1} is exponentially small there. The integral over the small circular arc is $O(n^{-\alpha-1})$ because the integrand is $O(n^{-\alpha})$ and the arc has length $O(n^{-1})$. On the two radial line segments, the modulus of z^{-n-1} is small except in a neighborhood of 1 of size $O(1/n)$. Here, changing variables to $(z-1)/n$, one has an integral of the form $\int A(z)\exp(n\phi(z))\,dz$, from which one obtains (e.g., via the well-known Watson–Doetsch lemma [39, Theorem 11.5]) the value

$$A(1)\,\frac{n^{-\alpha-1}}{\Gamma(-\alpha)}\,. \qquad \square \tag{2.4}$$

EXAMPLE 2.6 (Catalan numbers). *Let $a_n := \frac{1}{n+1}\binom{2n}{n}$ be the nth Catalan number. A great many naturally occurring combinatorial classes are counted by this sequence; there is a list of* 66 *of these in* [67, Problem 6.19]. *The generating function for the Catalan numbers is*

$$f(z) := \sum_{n=0}^{\infty} a_n z^n = \frac{1-\sqrt{1-4z}}{2z} = \frac{1-2\sqrt{\frac{1}{4}-z}}{2z}\,.$$

There is an algebraic singularity at $r = 1/4$, near which the asymptotic expansion for f begins

$$f(z) = 2 - 4\sqrt{\frac{1}{4}-z} + 8\left(\frac{1}{4}-z\right) - 16\left(\frac{1}{4}-z\right)^{3/2} + O\left(\frac{1}{4}-z\right)^{2}.$$

Note that $f/\sqrt{1/4-z}$ is not analytic in any disk of radius $1/4+\epsilon$, since both integral and half-integral powers appear in f, but f is analytic in a Camembert-shaped region. Theorem 2.5 *thus gives (note that the integral powers of $(1-z)$ do not contribute)*

$$\begin{aligned} a_n &\sim \left(\frac{1}{4}\right)^{1/2-n} n^{-3/2}\frac{-4}{\Gamma(-1/2)} + \left(\frac{1}{4}\right)^{3/2-n} n^{-5/2}\frac{-16}{\Gamma(-3/2)} + O(n^{-7/2}) \\ &= 4^n n^{-3/2}\frac{(-4)(1/4)^{1/2}}{\Gamma(-1/2)} + 4^n n^{-5/2}\frac{(-16)(1/4)^{3/2}}{\Gamma(-3/2)} + O(n^{-7/2}) \\ &= 4^n\left(\frac{n^{-3/2}}{\sqrt{\pi}} - n^{-5/2}\frac{3}{2\sqrt{\pi}} + O(n^{-7/2})\right). \end{aligned}$$

We have given only a brief view of what is known in the univariate case. We close this section with several more pointers to the literature.

One review of univariate asymptotics, which predates the work of Flajolet and Odlyzko but is still quite useful, is [6, Part II]. The 1995 survey article by Odlyzko [55], to which we have already referred, is somewhat more extensive. As mentioned before, contours such as the one in Figure 1 occur primarily in the univariate literature, but one recent extension to the multivariate setting, via a product of these contours, occurs in [26]: Lemma 3 in that paper gives a Darboux-type estimate for the $(n, k_1, \dots, k_j)$-coefficient of a generating function asymptotic to $(1-z_1)^{-\alpha} \prod_{i=1}^{j} (1-z_i)^{-\beta_i}$, uniform as long as $k_i = O(n)$ for all i. Finally, the book [21] is the most up to date, though it is only available in electronic preprint form at this time.

3. New Multivariate Results. In [57, 58, 4] the authors derived asymptotic formulae for $a_{\mathbf{r}}$ as $|\mathbf{r}| \to \infty$ that are uniform as $\mathbf{r}/|\mathbf{r}|$ varies over some compact set. It is useful to separate $\mathbf{r}$ into the scale parameter $|\mathbf{r}|$, which is a positive real number, and a direction parameter $\overline{\mathbf{r}}$, which is an element of real projective space. Although $\mathbf{r}$ is always an element of the positive orthant of $\mathbb{R}^d$, it will sometimes make sense to consider it as an element of $\mathbb{C}^d$. When convenient, we identify $\mathbf{r}$ with its class $\overline{\mathbf{r}}$ in projective space or its projection $\hat{\mathbf{r}}$ in the real $(d-1)$-simplex Δ^{d-1}, the set $\{\mathbf{x} \in (\mathbb{R}^+)^d : |\mathbf{x}| = 1\}$ (where $|\mathbf{x}| := \sum_{j=1}^{d} x_j$). Thus $\mathbf{r}$ may appear anywhere in the following diagram, where $\mathcal{O}$ and $\overline{\mathcal{O}}$ denote the positive orthants of $\mathbb{R}^d$ and $\mathbb{RP}^{d-1}$, respectively:

$$\begin{array}{ccccc} \mathcal{O} & \longrightarrow & \mathbb{R}^d & \longrightarrow & \mathbb{C}^d \\ {\scriptstyle -}\downarrow & & {\scriptstyle -}\downarrow & & {\scriptstyle -}\downarrow \\ \overline{\mathcal{O}} & \longrightarrow & \mathbb{RP}^{d-1} & \longrightarrow & \mathbb{CP}^{d-1} \\ {\scriptstyle \frown}\downarrow & & & & \\ \Delta^{d-1} & & & & \end{array}$$

Quantities that depend on $\mathbf{r}$ only through its direction will be denoted as functions of $\overline{\mathbf{r}}$. The results we will quote in this section may be informally summarized as follows. Recall that $F = G/H$.

(i) Asymptotics in the direction $\overline{\mathbf{r}}$ are determined by the geometry of the pole variety $\mathcal{V} = \{\mathbf{z} : H(\mathbf{z}) = 0\}$ near a finite set, $\mathsf{crit}_{\overline{\mathbf{r}}}$, of *critical points* (Definition 3.3).

(ii) For the purposes of asymptotic computation, one may reduce this set of critical points further to a set $\mathsf{contrib}_{\overline{\mathbf{r}}} \subseteq \mathsf{crit}_{\overline{\mathbf{r}}}$ of *contributing critical points*, usually a single point (formula (3.4) and Definition 3.4).

(iii) One may determine $\mathsf{crit}_{\overline{\mathbf{r}}}$ and $\mathsf{contrib}_{\overline{\mathbf{r}}}$ by a combination of algebraic and geometric criteria (Proposition 3.11 and Theorem 3.15, respectively).

(iv) Critical points may be of three types: smooth, multiple, or bad (Definition 3.9).

(v) Corresponding to each smooth or multiple critical point, $\mathbf{z}$, is an asymptotic expansion for $a_{\mathbf{r}}$ that is computable in terms of the derivatives of G and H at $\mathbf{z}$ (sections 3.3 and 3.4, respectively).

The culmination of the results above is the following metaformula:

$$a_{\mathbf{r}} \sim \sum_{\mathbf{z} \in \mathsf{contrib}_{\overline{\mathbf{r}}}} \mathbf{formula}(\mathbf{z}), \tag{3.1}$$

where $\mathbf{formula}(\mathbf{z})$ is one function of the local geometry for smooth points and a different function for multiple points. Specific instances of (3.1) are given in (3.6)–(3.13); the simplest case is

$$a_{\mathbf{r}} \sim C |\mathbf{r}|^{\frac{1-d}{2}} \mathbf{z}^{-\mathbf{r}},$$

where C and $\mathbf{z}$ are functions of $\overline{\mathbf{r}}$. No general expression for **formula** $(\mathbf{z})$ is yet known when $\mathbf{z}$ is a bad point, hence the name "bad point."

Fundamental to all the derivations is the Cauchy integral representation

$$a_{\mathbf{r}} = \frac{1}{(2\pi i)^d} \int_T \mathbf{z}^{-\mathbf{r}-\mathbf{1}} F(\mathbf{z})\, d\mathbf{z}, \tag{3.2}$$

where T is the product of sufficiently small circles around the origin in each of the coordinates, $\mathbf{1}$ is the d-vector of all ones, and $d\mathbf{z}$ is the holomorphic volume form $dz_1 \wedge \cdots \wedge dz_d$. The gist of the analyses in [57, 58, 4] is the computation of this integral. The key step is to replace the cycle T by a product $T_1 \times T_2$, where the inner integral over T_1 is a multivariate residue and the outer integral over T_2 is a saddle point integral. A complete asymptotic series may then be read off in a straightforward manner using well-known methods. The first subsection below explains what critical points and contributing critical points are and gives the "big picture," namely, the topological context as outlined in [4]. The next subsection gives the specific definitions from [57] that we need to find the contributing critical points (note: the question of sorting these into smooth, multiple, or bad critical points is addressed at the beginning of section 4). Sections 3.3 and 3.4 quote results from [57] and [58] that give asymptotics for smooth points if they are, respectively, smooth or multiple. Section 3.5 restates some of these asymptotics in terms of probability limit theorems.

3.1. Topological Representation. As $\mathbf{r} \to \infty$, the integrand in (3.2) becomes large. It is natural to attempt a saddle point analysis. That is, we try to deform the contour, T, so as to minimize the maximum modulus of the integrand. If $\hat{\mathbf{r}}$ remains fixed, then the modulus of the integrand is well approximated by the exponential term $\exp(-|\mathbf{r}|(\hat{\mathbf{r}} \cdot \log|\mathbf{z}|))$, where $\log|\mathbf{z}|$ is shorthand for the vector $(\log|z_1|, \ldots, \log|z_d|)$. This suggests that the real function

$$h(\mathbf{z}) := -\hat{\mathbf{r}} \cdot \log|\mathbf{z}| \tag{3.3}$$

be thought of as a *height function* in the Morse theoretic sense and that we try to deform T so that its maximum height is as low as possible. The stratified Morse theory [31] solves the problem of accomplishing such an optimal deformation. We now give a brief summary of this solution; for details, consult [4].

The variety $\mathcal{V}$ may be given a Whitney stratification. If $\mathcal{V}$ is already a manifold (which is the generic case), there is just a single stratum, but in general there may be any finite number of strata, even in two dimensions. The set of smooth points of $\mathcal{V}$ always constitutes the top stratum, with the set of singular points decomposing into the remaining strata, each stratum being a smooth manifold of lower complex dimension. To illustrate some of the possibilities, we give two examples; for further details on Whitney stratification, see [31, section I.1.2].

EXAMPLE 3.1 (complete normal intersection). *Let $\mathcal{V}$ be the variety where $(1-x)(1-y)(1-z)$ vanishes. The smooth points are those where exactly one of x, y, or z is equal to 1. There are three one-dimensional strata, where two but not all of $1-x, 1-y$, and $1-z$ vanish, and one zero-dimensional stratum at $(1,1,1)$.*

EXAMPLE 3.2 (isolated singular point). *Let $\mathcal{V}$ be the variety $\{1-x-y-z+4xyz = 0\}$. This has a zero-dimensional stratum at its unique singularity, $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$; everything else constitutes the two-dimensional stratum.*

In each stratum, there is a finite set of critical points, namely, points where the gradient of the height function restricted to the stratum vanishes. Since the height function depends on $\mathbf{r}$ via $\hat{\mathbf{r}}$, the set of critical points may be defined as a function of $\hat{\mathbf{r}}$ or, equivalently, of $\overline{\mathbf{r}}$.

DEFINITION 3.3 (critical points). *If S is a stratum of $\mathcal{V}$, define the set $\mathsf{crit}_{\overline{\mathbf{r}}}(S)$ of critical points of S for H in direction $\overline{\mathbf{r}}$ to be the set of $\mathbf{z} \in S$ for which $\nabla h|_S(\mathbf{z}) = 0$. The set $\mathsf{crit}_{\overline{\mathbf{r}}}$ of all critical points of $\mathcal{V}$ is defined as the union of $\mathsf{crit}_{\overline{\mathbf{r}}}(S)$ as S varies over all strata.*

If S has dimension k, then the condition $\nabla h|_S = 0$ defines an analytic variety of codimension k. Generically, then, $\mathsf{crit}_{\overline{\mathbf{r}}}(S)$ is zero-dimensional, that is, finite. It does happen sometimes that for a small set of values of $\overline{\mathbf{r}}$ the critical set is not finite; for example, this happens when $\mathcal{V}$ is a binomial variety $\{\mathbf{z} : \mathbf{z}^{\mathbf{a}} - \mathbf{z}^{\mathbf{b}} = 0\}$, in which case $\nabla h|_S$ vanishes everywhere on S for a particular $\overline{\mathbf{r}}$ and nowhere on S for other values of $\overline{\mathbf{r}}$. We address such a case in section 4.11. Otherwise, such a degeneracy does not occur among our examples and a standing assumption (Assumption 3.6 below) will ensure that it does not cause trouble in any case.

If F is a rational function, then the critical points are a finite union of zero-dimensional varieties, so elimination theory (see, e.g., [15, Chapters 1 and 2]) will provide, in an automatic way, minimal polynomials for the critical points; see Proposition 3.11 below. For future reference, we enumerate the critical points as $\{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}\}$.

Let $\mathcal{M}$ denote the domain of holomorphy of the integrand in (3.2); that is,

$$\mathcal{M} := \mathbb{C}^d \setminus \left\{ \mathbf{z} : H \cdot \prod_{j=1}^{d} z_j = 0 \right\} .$$

The topology of $\mathcal{M}$ is determined by neighborhoods of the critical points of $\mathcal{V}$ (the complement of $\mathcal{M}$) in a manner we now describe. Given a real number c, let $\mathcal{M}^c$ denote the set of points of $\mathcal{M}$ with height less than c. Let $\mathcal{M}^+$ denote the $\mathcal{M}^c$ for some c greater than the greatest height of any critical point and let $\mathcal{M}^-$ denote $\mathcal{M}^c$ for some c less than the least height of any critical point. If the interval $[c, c']$ does not contain a critical value of the height function, then $\mathcal{M}^c$ is a strong deformation retract of $\mathcal{M}^{c'}$, so the particular choices of c above do not matter to the topology and hence to the evaluation of the integral (3.2).

Let X denote the topological pair $(\mathcal{M}^+, \mathcal{M}^-)$. The homology group $H_d(X)$ is generated by the homology groups $H_d(\mathcal{M}^{c'}, \mathcal{M}^{c''})$ as $[c', c'']$ ranges over a finite set of intervals whose union contains all critical values. These groups in turn are generated by quasi-local cycles. At any critical point $\mathbf{z}^{(i)}$ in the top stratum, the quasi-local cycle is the Cartesian product of a patch $\mathcal{P}_i$ diffeomorphic to $\mathbb{R}^{d-1}$ inside $\mathcal{V}$, whose maximum height is achieved at $\mathbf{z}^{(i)}$, with an arbitrarily small circle γ_i transverse to $\mathcal{V}$.

Thus, for example, when $d = 2$, the quasi-local cycles near smooth points look like pieces of macaroni: a product of a circle with a rainbow-shaped arc whose peak is at $\mathbf{z}^{(i)}$ (See Figure 2). We have not described the quasi-local cycles for nonmaximal strata, but for a description of the quasi-local cycles in general, one may look in [4] or [31].[1]

[1]For generic $\overline{\mathbf{r}}$, the function h is Morse. Each attachment map is a d-dimensional complex, so the attachments induce injections on $H_d(X)$. It may happen for some $\overline{\mathbf{r}}$ that h is not Morse, but in this case one may understand the local topology via a small generic perturbation. Two or more critical points may merge, but the fact that the attachments induce injections on $H_d(X)$ implies that such a merger produces a direct sum in $H_d(X)$; in particular, it will be useful later to know that a cycle in the merger is nonzero if and only if at least one component is nonzero.

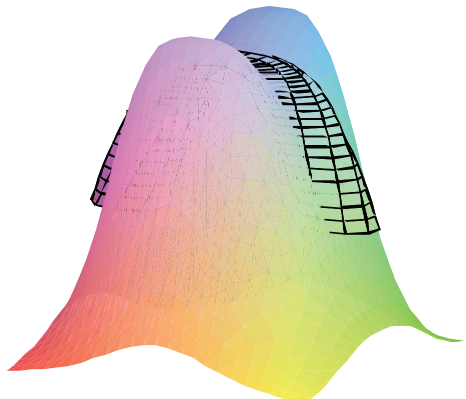

Fig. 2 *A piece of a quasi-local cycle at top-dimensional critical point.*

Since the quasi-local cycles generate $H_d(\mathcal{M}^+, \mathcal{M}^-)$, it follows that the integral (3.2) is equal to an integer linear combination

$$\sum n_i \int_{\mathcal{C}_i} \mathbf{z}^{-\mathbf{r}-\mathbf{1}} F(\mathbf{z})\, d\mathbf{z}, \tag{3.4}$$

where $\mathcal{C}_i$ is a quasi-local cycle near $\mathbf{z}$ and $T = \sum n_i \mathcal{C}_i$ in $H_d(\mathcal{M}^+, \mathcal{M}^-)$. When $\mathbf{z}$ is in the top stratum, i.e., $\mathbf{z}$ is a smooth point of $\mathcal{V}$, then $\mathcal{C}_i = \mathcal{P}_i \times \gamma_i$, the product of a d-patch with a transverse circle, so the integral may be written as $\int_{\mathcal{P}_i} \int_{\gamma_i} \exp[-|\mathbf{r}| h(\mathbf{z})] F(\mathbf{z})\, d\mathbf{z}$. The inner integral is a simple residue and the outer integral is a standard saddle point integral. The asymptotic evaluation of $a_{\mathbf{r}}$ is therefore solved if we can compute the integers n_i in the decomposition of T into $\sum n_i \mathcal{C}_i$. Observe that the contribution from $\mathbf{z}^{(i)}$ is of exponential order no greater than $\exp(h(\mathbf{z}^{(i)}))$. It turns out (Theorem 3.19 and the formulae of section 3.4) that for smooth and multiple points, this is a lower bound on the exponential order as well, provided, in the multiple point case, that $G(\mathbf{z}^{(i)}) \neq 0$. Thus if $n_i \neq 0$, one may ignore any contributions from $\mathbf{z}^{(j)}$ with $h(\mathbf{z}^{(j)}) < h(\mathbf{z}^{(i)})$ and still obtain an asymptotic expansion containing all terms not exponentially smaller than the leading term.

Definition 3.4 (contributing critical points). *The set* $\mathsf{contrib} = \mathsf{contrib}_{\overline{\mathbf{r}}}$ *of contributing critical points is defined to be the set of* $\mathbf{z}^{(i)}$ *such that* $n_i \neq 0$ *and* $n_j = 0$ *for all* j *with* $h(\mathbf{z}^{(j)}) > h(\mathbf{z}^{(i)})$. *In other words, the contributing critical point(s) are just the highest ones with nonzero coefficients,* n_i. *Note that if there is more than one contributing critical point, all must have the same height.*

The problem of computing the topological decomposition may be difficult in general and no complete solution is known. In two dimensions, in the case where $\mathcal{V}$ is globally smooth, an effective algorithm is known for finding the contributing critical points. The algorithm follows approximate steepest descent paths; the details have not yet been published [69]. In any dimension, if the variety $\mathcal{V}$ is the union of hyperplanes, it is shown in [4] how to evaluate all the n_i; see also some preliminary work on this in [10]. In the case where $\mathcal{V}$ is not smooth, while no general solution is known, we may still state a sufficient condition for the critical point $\mathbf{z}^{(i)}$ to be a contributing critical point. It is to these geometric sufficient conditions that we turn next.

3.2. Geometric Criteria. Let F, G, H, and $a_{\mathbf{r}}$ be as in (1.3). If $a_{\mathbf{r}} \geq 0$ for all $\mathbf{r}$, we say we are in the *combinatorial case*. We assume throughout that G and H are relatively prime in the ring of analytic functions on the domain of convergence of F. We wish to compute the function contrib which maps directions in $\overline{\mathcal{O}}$ to finite subsets of $\mathcal{V}$ (often singletons). The importance of *minimal points*, defined below, is that when $\mathsf{contrib}_{\hat{\mathbf{r}}}$ contains a minimal point, this point may be identified by a variational principle.

DEFINITION 3.5 (notation for polydisks and tori). *Let* $\mathbf{D}(\mathbf{z})$ *and* $\mathbf{T}(\mathbf{z})$ *denote, respectively, the polydisk and torus defined by*

$$\mathbf{D}(\mathbf{z}) := \{\mathbf{z}' : |z_i'| \le |z_i| \text{ for all } 1 \le i \le d\},$$

$$\mathbf{T}(\mathbf{z}) := \{\mathbf{z}' : |z_i'| = |z_i| \text{ for all } 1 \le i \le d\}.$$

The open domain of convergence of F is denoted $\mathcal{D}$ and is the union of tori $\mathbf{T}(\mathbf{z})$. The logarithmic domain of convergence, namely, those $\mathbf{x} \in \mathbb{R}^d$ with $(e^{x_1}, \dots, e^{x_d}) \in \mathcal{D}$, is denoted $\log\mathcal{D}$ *and is always convex* [40]. *The image $\{\log|\mathbf{z}| : \mathbf{z} \in \mathcal{V}\}$ of $\mathcal{V}$ under the coordinatewise log-modulus map is denoted $\log\mathcal{V}$ (this is sometimes called the* amoeba *of $\mathcal{V}$* [28]).

ASSUMPTION 3.6. *We assume throughout that* $\log\mathcal{D}$ *is strictly convex, that is, its boundary contains no line segment.*

This assumption ensures that $h_{\hat{\mathbf{r}}}$ is always uniquely minimized on $\partial\log\mathcal{D}$. The assumption is satisfied in all examples in this paper. It is a consequence of $\log\mathcal{V}$ containing no line segments, which is true in all examples outside of section 4.11. This may be checked by computer algebra; we do not include the checking of this fact in our worked examples.

DEFINITION 3.7 (minimality). *A point $\mathbf{z} \in \mathcal{V}$ is* minimal *if all coordinates are nonzero and the relative interior of the associated polydisk contains no element of $\mathcal{V}$, that is, $\mathcal{V} \cap \mathbf{D}(\mathbf{z}) \subseteq \partial\mathbf{D}(\mathbf{z})$. More explicitly, $z_i \neq 0$ for all i and there is no $\mathbf{z}' \in \mathcal{V}$ with $|z_i'| < |z_i|$ for all i. The minimal point $\mathbf{z}$ is said to be* strictly minimal *if it is the only point of $\mathcal{V}$ in the closed polydisk: $\mathcal{V} \cap \mathbf{D}(\mathbf{z}) = \{\mathbf{z}\}$.*

Remark. This definition is equivalent to the apparently stronger definition of minimality stated in [57], namely, that $\mathcal{V} \cap \mathbf{D}(\mathbf{z}) \subseteq \mathbf{T}(\mathbf{z})$; see Proposition 3.12.

PROPOSITION 3.8 (minimal points in contrib). *Recall Assumption* 3.6.

(i) *A point $\mathbf{z} \in \mathcal{V}$ with nonzero coordinates is minimal if and only if $\mathbf{z} \in \partial\mathcal{D}$.*

(ii) *All minimal points in $\mathsf{contrib}_{\overline{\mathbf{r}}}$ must lie on the same torus and the height $h_{\hat{\mathbf{r}}}(\mathbf{z})$ of the point(s) in $\mathsf{contrib}_{\overline{\mathbf{r}}}$ is (are) at most the minimum, c, of the height function $h_{\mathbf{r}}$ on $\partial\mathcal{D}$.*

(iii) *If $\mathbf{z} \in \mathsf{contrib}_{\overline{\mathbf{r}}} \cap \partial\mathcal{D}$, then the hyperplane normal to $\overline{\mathbf{r}}$ at $\log|\mathbf{z}|$ is a support hyperplane to* $\log\mathcal{D}$.

To summarize, if $\mathsf{contrib}_{\overline{\mathbf{r}}}$ contains any minimal points, then these points minimize $h_{\mathbf{r}}$ on $\mathcal{D}$ and they all project, via coordinatewise log-modulus, to a single point of $\partial\log\mathcal{D}$, where the support hyperplane is normal to $\mathbf{r}$.

Proof. A power series converges absolutely on any open polydisk about the origin in which it is analytic; this can be seen by using Cauchy's formula (3.2) to estimate $a_{\mathbf{r}}$. Thus any minimal point is in the closure of $\mathcal{D}$. A series cannot converge at a pole, so $\mathbf{z} \notin \mathcal{D}$ and the first assertion follows.

To prove the second assertion, by strict convexity there is a unique $\mathbf{z} \in \partial\mathcal{D}$ that minimizes h. Let $\mathbf{H}$ be the homotopy $\{t\mathbf{T}(\mathbf{z}) : \epsilon \le t < 1 - \epsilon\}$. This may be extended to a homotopy $\mathbf{H}'$ that pushes the height below $c - \epsilon$ except in small neighborhoods of $\mathcal{V} \cap \mathbf{T}(\mathbf{z})$. Thus the small torus T in (3.2) is homotopic to a cycle in the union of $\mathcal{M}^{c-\epsilon}$ and a neighborhood of $\mathcal{V} \cap \mathbf{T}(\mathbf{z})$. Letting $\epsilon \to 0$, it follows that contrib has height at most c. If there is a line segment in $\partial\log\mathcal{D}$, it is possible that the minimum height on

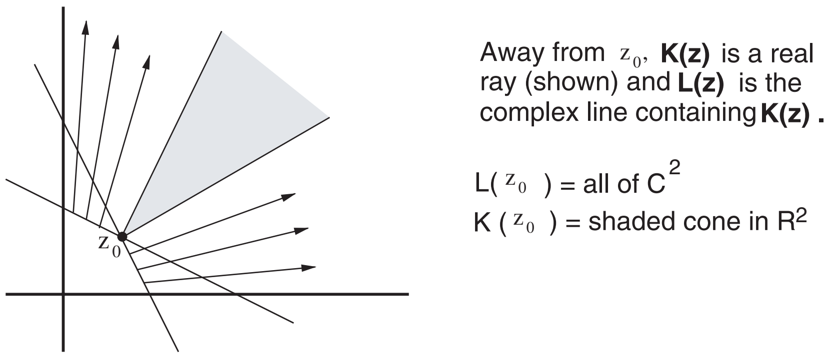

Fig. 3 *Direction* $\mathbf{L}(\mathbf{z})$ *for positive real points of* $\mathcal{V}$ *when* $1/F = (3-x-2y)(3-2x-y)$.

$\mathcal{D}$ occurs on more than one torus, but in this case, if $\mathbf{z}$ and $\mathbf{w}$ are on different such tori, the above argument shows that neither $\mathbf{z}$ nor $\mathbf{w}$ can be in $\mathsf{contrib}_{\overline{\mathbf{r}}}$.

The third assertion is immediate from the fact that $\log|\mathbf{z}|$ minimizes the linear function $h_{\hat{\mathbf{r}}}$ on the convex set $\log\mathcal{D}$. ∎

Having characterized $\mathsf{contrib}_{\overline{\mathbf{r}}}$ when it is in $\partial\mathcal{D}$, via a variational principle, we now relate this to the algebraic definition of $\mathsf{crit}_{\overline{\mathbf{r}}}$, which is better for symbolic computation. It is easy to write down the inverse of $\mathsf{crit}_{\overline{\mathbf{r}}}$, which we will denote by $\mathbf{L}$.

Definition 3.9 (geometry of minimal points). $\mathbf{z}$ *is a* smooth point *of* $\mathcal{V}$ *if* $\mathcal{V}$ *is a manifold in a neighborhood of* $\mathbf{z}$*.* $\mathbf{z}$ *is a* multiple point *of* $\mathcal{V}$ *if* $\mathcal{V}$ *is the union of finitely many manifolds near* $\mathbf{z}$ *intersecting transversely (that is, normals to any* k *of these manifolds span a space of dimension* $\min\{k,d\}$*). Points that are neither smooth nor multiple are informally called bad points.*

Remark. The transversality assumption streamlines the arguments but is not really needed (see, e.g., [4]).

Definition 3.10 (the linear space $\mathbf{L}$). *Let* $\mathbf{z} \in \mathcal{V}$ *be in a stratum* $\mathcal{S}$*. Let* $\mathbf{L}(\mathbf{z}) \subseteq \mathbb{CP}^{d-1}$ *denote the span of the projections of vectors* $(z_1 v_1, \ldots, z_d v_d)$ *as* $\mathbf{v}$ *ranges over vectors orthogonal to the tangent space of* $\mathcal{S}$ *at* $\mathbf{z}$*.*

Figure 3 gives a pictorial example of the previous definitions (some parts of the figure refer to Definition 3.13 below). The singular variety $\mathcal{V}$ for the function $F(x,y) = 1/[(3-x-2y)(3-2x-y)]$ is composed of two complex lines meeting at $(1,1)$. The crossing point $\mathbf{z}_0 := (1,1)$ is a singleton stratum. The following proposition, showing why $\mathbf{L}$ is important, follows directly from the definitions.

Proposition 3.11 ($\mathbf{L}$ inverts crit). *The point* $\mathbf{z}$ *is in* $\mathsf{crit}_{\overline{\mathbf{r}}}$ *if and only if* $\overline{\mathbf{r}} \in \mathbf{L}(\mathbf{z})$*. If* $\mathbf{z}$ *is a smooth point, then (replacing* H *by its radical if needed)* $\mathbf{L}(\mathbf{z})$ *is the singleton set (in projective space)*

$$\mathbf{L}(\mathbf{z}) = \{\nabla_{\log} H(\mathbf{z})\} := \left\{ \left(z_1 \frac{\partial H}{\partial z_1}(\mathbf{z}), \ldots, z_d \frac{\partial H}{\partial z_d}(\mathbf{z}) \right) \right\},$$

while the set $\mathsf{crit}_{\overline{\mathbf{r}}}$ *is the (usually zero-dimensional) variety given by the* d *equations*

$$H = 0,$$

$$r_d z_j \frac{\partial H}{\partial z_j} = r_j z_d \frac{\partial H}{\partial z_d} \qquad (1 \le j \le d-1). \tag{3.5}$$

If $\mathbf{z}$ *is a multiple point, then* $\mathbf{L}(\mathbf{z})$ *is the span of the vectors*

$$\nabla_{\log} H_k(\mathbf{z}) := \left(z_1 \frac{\partial H_k}{\partial z_1}(\mathbf{z}), \ldots, z_d \frac{\partial H_k}{\partial z_d}(\mathbf{z}) \right),$$

where $H = \prod_k H_k$. ◻

Connecting this with the variational principle, we have the following proposition.

PROPOSITION 3.12 (smooth minimal points are critical). *If* $\mathbf{z}$ *is minimal and smooth, then* $\mathbf{L}(\mathbf{z}) \in \overline{\mathcal{O}}$ *and is normal to a support hyperplane to* $\log \mathcal{D}$ *at* $\log|\mathbf{z}|$. *Consequently, a minimal smooth point* $\mathbf{z}$ *is a critical point for some outward normal direction to* $\log \mathcal{D}$ *at* $\log|\mathbf{z}|$.

Proof. Assume first that none of the logarithmic partial derivatives $h_j := z_j \partial H/\partial z_j$ vanishes at $\mathbf{z}$. Suppose the arguments of two of the partial derivatives h_k and h_l are not equal. Since we have assumed no h_j vanishes, there is a tangent vector $\mathbf{v}$ to $\mathcal{V}$ with $v_j = 0$ for $j \notin \{k, l\}$ and $v_k \neq 0 \neq v_l$. We may choose a multiple of this so that $\overline{v_k} h_k$ and $\overline{v_l} h_l$ both have negative real parts. Perturbing slightly, we may choose a tangent vector $\mathbf{v}$ to $\mathcal{V}$ at $\mathbf{z}$ with all nonzero coordinates so that $\overline{v_j} h_j$ has negative real part for all j. This implies there is a path from $\mathbf{z}$ on $\mathcal{V}$ such that the moduli of all coordinates strictly decrease, which contradicts minimality. It follows that all arguments of h_j are equal, and therefore that $\mathbf{L}(\mathbf{z})$ is a point of $\overline{\mathcal{O}} \subseteq \mathbb{CP}^d$; hence it is normal to a support hyperplane to $\log \mathcal{D}$ at $\log|\mathbf{z}|$.

Now, suppose $h_j = 0$ for j in some nonempty set, J. Let $\mathbf{x} = \log|\mathbf{z}|$. Vary $\{z_j : j \notin J\}$ by $\Theta(\epsilon)$ so that it stays on $\mathcal{V}$, varying $\{z_j : j \in J\}$ by $O(\epsilon^2)$. We see that the complement of $\log \mathcal{D}$ contains planes arbitrarily close to the $|J|$-dimensional real plane through $\mathbf{x}$ in directions $\{e_j : j \in J\}$, whence the closure of the complement contains this plane. But, by monotonicity, the orthant of this plane in the $-e_j$ direction must be in the closure of $\log \mathcal{D}$. Thus this orthant is in $\partial \log \mathcal{D}$, whence there is a lifting of this orthant to $\mathcal{V}$. Now the argument by contradiction in the previous paragraph shows that if the arguments of h_k and h_l are unequal for some $k, l \notin J$, then we may move on $\mathcal{V}$ so the moduli of the z_j for $j \notin J$ strictly decrease, while moving down the lifting of the orthant allows us also to decrease the moduli of the z_j for $j \in J$, and we have again contradicted minimality. ◻

Having related critical points to $\mathbf{L}(\mathbf{z}) \subseteq \mathbb{CP}^{d-1}$, we now relate contributing critical points to a set $\mathbf{K}(\mathbf{z}) \subseteq \mathbb{RP}^{d-1}$. An open problem is to find the right definition of $\mathbf{K}(\mathbf{z})$ when $\mathbf{z}$ is not minimal.

DEFINITION 3.13 (the cone $\mathbf{K}$ of a minimal point). *If* $\mathbf{z}$ *is a smooth point of* $\mathcal{V}$ *in* $\partial \mathcal{D}$, *let* $\mathsf{dir}(\mathbf{z})$ *be defined as* $\mathbf{L}(\mathbf{z})$, *which is in* $\overline{\mathcal{O}}$ *by Proposition* 3.12. *For minimal* $\mathbf{z} \in \mathcal{V}$, *not necessarily smooth, define* $\mathbf{K}(\mathbf{z})$ *to be the convex hull of the set of limit points of* $\mathsf{dir}(\mathbf{w})$ *as* $\mathbf{w} \to \mathbf{z}$ *through smooth points.*

For an example of this, see Figure 3, in which arrows are drawn depicting $\mathbf{K}(\mathbf{z})$ for smooth points $\mathbf{z}$ in the positive real quadrant and the two-dimensional cone $\mathbf{K}(\mathbf{z}_0)$ is shaded. The next proposition then follows from Proposition 3.11 and basic properties of convex sets.

PROPOSITION 3.14 (description of $\mathbf{K}$). *Let* $\mathbf{z}$ *be a minimal point.*

(i) $\mathbf{K}(\mathbf{z}) \subseteq \mathbf{L}(\mathbf{z}) \cap \overline{\mathcal{O}}$.

(ii) *If* $\mathbf{z}$ *is a smooth or multiple point, then* $\mathbf{K}(\mathbf{z})$ *is the cone spanned by the vectors* $\nabla_{\log} H_k$ *of Proposition* 3.11.

(iii) *If a neighborhood of* $\mathbf{z}$ *in* $\mathcal{V}$ *covers a neighborhood of* $\log|\mathbf{z}|$ *in* $\partial \log \mathcal{D}$, *then* $\mathbf{K}(\mathbf{z})$ *is the set of outward normals to* $\log \mathcal{D}$ *at* $\log|\mathbf{z}|$.

The first nontrivial result we need is the following theorem.

THEOREM 3.15 ($\mathbf{K}$ inverts $\mathsf{contrib}$). *Let* $\mathbf{z}$ *be a minimal point that is either smooth or multiple. Then* $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$ *if and only if* $\mathbf{z} \in \mathsf{contrib}_{\overline{\mathbf{r}}}$*, provided, in the multiple point case, that* $G(\mathbf{z}) \neq 0$.

Remark. Here, rather than global meromorphicity, it is only required that F be meromorphic in a neighborhood of $\mathbf{D}(\mathbf{z})$; see [57] and [58] for details.

Proof. Assume first that $\mathbf{z}$ is strictly minimal, that is, $\mathbf{T}(\mathbf{z}) \cap \mathcal{V} = \{\mathbf{z}\}$.

Suppose for now that $\overline{\mathbf{r}}$ is in the relative interior of $\mathbf{K}(\mathbf{z})$. Theorems 3.5 of [57] and 3.6 and 3.9 of [58] give expressions for $a_{\mathbf{r}}$ which are of the order $|\mathbf{r}|^{\beta}\mathbf{z}^{-\mathbf{r}}$, relying, in the multiple point case, on the assumption of transverse intersection, which we have built into our definition of multiple point, and on $G(\mathbf{z}) \neq 0$.

On one hand, it is evident from (3.2) that

$$a_{\mathbf{r}} = O\left(\exp\left[|\mathbf{r}|(h_{\hat{\mathbf{r}}}(\mathbf{w}) + \epsilon)\right]\right)$$

for any $\epsilon > 0$, where $\mathbf{w} \in \mathsf{contrib}_{\overline{\mathbf{r}}}$. On the other hand, we know by Proposition 3.8 that no points higher than $\mathbf{z}$ are in $\mathsf{contrib}_{\overline{\mathbf{r}}}$. Since $\mathsf{contrib}_{\overline{\mathbf{r}}}$ is nonempty, we conclude that it has points at height $h_{\hat{\mathbf{r}}}(\mathbf{z})$, whence from the expressions for $a_{\mathbf{r}}$ again, we see that $\mathbf{z} \in \mathsf{contrib}_{\overline{\mathbf{r}}}$.

In the case $\overline{\mathbf{r}} \in \partial\mathbf{K}(\mathbf{z})$, consider $\overline{\mathbf{r}}' \to \overline{\mathbf{r}}$ through the interior of $\mathbf{K}(\mathbf{z})$ (here we mean relative boundary and relative interior). There will be a contributing point $\mathbf{z}'$ which converges to $\mathbf{z}$. The coefficient of the quasi-local cycle at $\mathbf{z}$ in the limit must be nonzero (see footnote 1). The theorem is now proven for strictly minimal points.

To remove the assumption of strict minimality, we must verify that this was not necessary for the formulae we quoted from [57, 58]. These formulae were proved by reducing the Cauchy integral to an integral over a neighborhood $\mathcal{N}$ of a $(d-1)$-dimensional subset Θ of $\mathcal{V}$. It is pointed out in [57, Corollary 3.7] that it is sufficient to assume *finite minimality*, that is, finiteness of $\mathbf{T}(\mathbf{z}) \cap \mathcal{V}$. In fact, one needs only finiteness of $\mathbf{T}(\mathbf{z}) \cap \Theta$, since the truncation of the integral to $\mathcal{N}$ incurs a boundary term that is sufficiently small as long as Θ avoids $\mathbf{T}(\mathbf{z})$. The remaining case, where $\Theta \cap \mathbf{T}(\mathbf{z})$ is infinite, can be handled by contour rotation arguments, but since that work is not yet published, we point out here that in the case $d = 2$ (the only case used in this survey), the set Θ is one-dimensional. Thus, when $\Theta \cap \mathbf{T}(\mathbf{z})$ is infinite, the set Θ must be a subset of $\mathbf{T}(\mathbf{z})$; the integrals in [57, 58] may then be taken over all of Θ with no truncation. ∎

We remarked earlier that the main challenge in computing asymptotics is to identify $\mathsf{contrib}_{\overline{\mathbf{r}}}$. Our progress to this point is that we may find all strictly minimal smooth or multiple points in $\mathsf{contrib}_{\overline{\mathbf{r}}}$ by solving the equation $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$ for $\mathbf{z}$. This equation may be solved automatically when F is a rational function and is often tractable in other cases, for example, in the case $F = (e^x - e^y)/(xe^y - ye^x)$ of section 4.7.

To complement this, we would like to know when solving $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$ and checking for minimality does indeed find all points of $\mathsf{contrib}_{\overline{\mathbf{r}}}$. It cannot, for instance, do so if there are no minimal contributing points. Thus we ask, (a) are there any minimal points $\mathbf{z}$ with $\mathbf{K}(\mathbf{z}) = \overline{\mathbf{r}}$; (b) are all the contributing points minimal; and (c) is there more than one minimal point? It turns out that the answers are, roughly, (a) yes, in the combinatorial case; (b) yes, as long as (a) is true, by part (ii) of Proposition 3.8; and (c) rarely (we know they are never on different tori).

Let $\boldsymbol{\Xi} \subseteq \overline{\mathcal{O}}$ denote the set of all normals to support hyperplanes of $\log\mathcal{D}$. If the restriction of F to each coordinate hyperplane is not entire, then $\boldsymbol{\Xi}$ is the entire nonnegative orthant. This is because $\log\mathcal{D}$ has support hyperplanes parallel to each coordinate hyperplane. When $\boldsymbol{\Xi}$ is not the whole orthant, then, in directions $\overline{\mathbf{r}} \notin \overline{\mathcal{O}}$, the quantity $a_{\mathbf{r}}$ is either identically zero or decays faster than exponentially (this follows, for example, from [58, equation (1.2)], since $-\hat{\mathbf{r}} \cdot \log|\mathbf{z}|$ is not bounded from below on $\mathcal{D}$).

THEOREM 3.16 (existence of minimal points in the combinatorial case). *Suppose that we are in the combinatorial case, $a_{\mathbf{r}} \geq 0$. Under the standing assumption* 3.6*, for every $\overline{\mathbf{r}}$ in the interior of Ξ, there is a minimal point $\mathbf{z} \in \mathcal{V}$ which lies in the positive orthant of $\mathbb{R}^d$ and has $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$.*

Remark. At this point, it is clear how far one can generalize beyond rational functions. Given a compact set $K \subseteq \Xi$ in which one wishes to find asymptotics, let $\log\mathcal{D}(K)$ denote the set of all $\mathbf{x}$ such that $\mathbf{x} \leq \mathbf{y}$ coordinatewise for some $\mathbf{y} \in \partial\log\mathcal{D}$ whose normal direction is in K. Let $\mathcal{D}(K)$ be the inverse image under the coordinatewise log-modulus map of $\log\mathcal{D}(K)$. Then, in the combinatorial case, for the existence of a minimal point $\mathbf{z}$ in the positive orthant with $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$, it is sufficient that F be meromorphic in a neighborhood of $\mathcal{D}(K)$.

Proof. We follow the proof of [57, Theorem 6.3]. For any $\overline{\mathbf{r}}$ in the interior of Ξ, there is a point $\mathbf{x} \in \partial\log\mathcal{D}$ with a support hyperplane normal to $\overline{\mathbf{r}}$. Let $z_i = e^{x_i}$ so $\mathbf{z}$ is a real point in $\partial\mathcal{D}$. We claim that $\mathbf{z} \in \mathcal{V}$. To see this, note that there is a singularity on $\mathbf{T}(\mathbf{z})$, which must be a pole by the assumption of meromorphicity on a neighborhood of $\mathbf{D}(\mathbf{z})$. Together, $a_{\mathbf{r}} \geq 0$ and lack of absolute convergence of the series on $\mathbf{T}(\mathbf{z})$ imply that $F(\mathbf{w})$ converges to $+\infty$ as $\mathbf{w} \to \mathbf{z}$ from beneath. By meromorphicity, $\mathbf{z}$ is therefore a pole of F, so $\mathbf{z} \in \mathcal{V}$.

We conclude that there is a lifting of $\partial\log\mathcal{D}$ to the real points of $\mathcal{V}$, that is, a subset of the real points of $\mathcal{V}$ maps properly and one to one onto $\partial\log\mathcal{D}$. By the last part of Proposition 3.14, it follows that $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$, and hence from Theorem 3.15 that $\mathbf{z} \in \mathsf{contrib}_{\overline{\mathbf{r}}}$. ∎

Remark. By convexity of $\log\mathcal{D}$, the point $\mathbf{z} \in \overline{\mathcal{O}}^d \cap \mathsf{crit}(\overline{\mathbf{r}})$ in Theorem 3.16 is unique unless there is a line segment in $\partial\log\mathcal{D}$. In this case there is a continuum of such $\mathbf{z}$. Since in almost every example the critical point variety is finite, there will be precisely one critical point in the positive orthant. Thus in practice we will have no difficulty in determining the point $\mathbf{z}$.

We say that a power series P is *aperiodic* if the sublattice of $\mathbb{Z}^d$ of integer combinations of exponent vectors of the monomials of P is all of $\mathbb{Z}^d$. By a change of variables, we lose no generality from the point of view of generating functions if we assume in the following proposition that P is aperiodic.

PROPOSITION 3.17 (often, every minimal point is strictly minimal). *If $H = 1-P$, where P is aperiodic with nonnegative coefficients, then every minimal point is strictly minimal and lies in the positive orthant.*

Proof. Suppose that $\mathbf{z} = \mathbf{x}e^{i\theta}$, with $\theta \neq \mathbf{0}$, is minimal. Since $\mathbf{z} \in \mathcal{V}$, we have

$$\begin{aligned} 1 &= \left| \sum_{\mathbf{r} \neq \mathbf{0}} a_{\mathbf{r}} \mathbf{x}^{\mathbf{r}} \exp(i\mathbf{r} \cdot \theta) \right| \\ &\leq \sum_{\mathbf{r} \neq \mathbf{0}} a_{\mathbf{r}} \mathbf{x}^{\mathbf{r}} . \end{aligned}$$

Thus $\mathbf{z}$ can be strictly minimal only if $\theta = \mathbf{0}$.

Equality holds in the above inequality if and only if either there is only one term in the latter sum, or, when there is more than one such term, $\exp(i\mathbf{r} \cdot \theta) = 1$ whenever $a_{\mathbf{r}} \neq 0$. The former case occurs precisely when P has the form $c\mathbf{x}^{\mathbf{r}}$ and the latter can occur when P has the form $P(\mathbf{z}) = g(\mathbf{z}^{\mathbf{b}})$ for some $\mathbf{b} \neq 0$, both of which are ruled out by aperiodicity. ∎

We restate the gist of Theorem 3.16 and Proposition 3.17 as the following corollary.

COROLLARY 3.18. *If $H = 1 - P$ with P aperiodic and with nonnegative coefficients, then the contributing critical points as $\bar{\mathbf{r}}$ varies are precisely the points of $\mathcal{V} \cap \mathcal{O}^d$ that are minimal in the coordinatewise partial order. The point $\mathbf{z}$ is in $\mathsf{contrib}_{\bar{\mathbf{r}}}$ exactly when $\bar{\mathbf{r}} \in \mathbf{K}(\mathbf{z})$.*

Without the assumption $H = 1 - P$, assuming only $a_{\mathbf{r}} \geq 0$, the same holds except that there might be more contributing critical points on the torus $\mathbf{T}(\mathbf{z})$. □

3.3. Asymptotic Formulae for Minimal Smooth Points. It is shown in [57] how to compute the integral (3.2) when $\bar{\mathbf{r}}$ is fixed and $\mathsf{contrib}_{\bar{\mathbf{r}}}$ is a finite set of smooth points on a torus. Given a smooth point $\mathbf{z} \in \mathcal{V}$, let $\tilde{f}_{\mathbf{z}}$ be the map on a neighborhood of the origin in $\mathbb{R}^{d-1}$ taking the origin to zero and taking $(\theta_1, \ldots, \theta_{d-1})$ to $\log w$ for w such that

$$\left(z_1 e^{i\theta_1}, \ldots, z_{d-1} e^{i\theta_{d-1}}, z_d w\right) \in \mathcal{V}.$$

The following somewhat general result is shown in [57, Theorem 3.5].

THEOREM 3.19 (smooth point asymptotics). *Let $K \subset \Xi$ be compact, and suppose that for $\bar{\mathbf{r}} \in K$, the set $\mathsf{contrib}_{\bar{\mathbf{r}}}$ is a single smooth point $\mathbf{z}(\bar{\mathbf{r}})$ and $\tilde{f}_{\mathbf{z}}$ has nonsingular Hessian (matrix of second partial derivatives). Then there are effectively computable functions $b_l(\bar{\mathbf{r}})$ such that*

$$a_{\mathbf{r}} \sim \mathbf{z}(\bar{\mathbf{r}})^{-\mathbf{r}} \sum_{l \geq 0} b_l(\bar{\mathbf{r}}) (r_d)^{-(l+d-1)/2} \tag{3.6}$$

as an asymptotic expansion when $|\mathbf{r}| \to \infty$, uniformly for $\bar{\mathbf{r}} \in K$. □

Remark. In connection with Theorem 3.19, the following points should be noted.

(i) The coefficients b_l depend on the derivatives of G and H to order $k + l - 1$ at $\mathbf{z}(\bar{\mathbf{r}})$, and $b_0(\bar{\mathbf{r}}) = 0$ if and only if $G(\mathbf{z}(\bar{\mathbf{r}})) = 0$.

(ii) The sum on the right of (3.6) may be rewritten as a sum of the terms $b_l^*(\bar{\mathbf{r}})|\mathbf{r}|^{-(l+d-1)/2}$, where $b_l^* = \hat{\mathbf{r}}_d^{-(l+d-1)/2} b_l$, in order to see what depends on $\hat{\mathbf{r}}$ and what depends on $|\mathbf{r}|$.

(iii) Suppose $\mathsf{contrib}_{\bar{\mathbf{r}}}$ is a finite set of points $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(k)}$, satisfying the hypotheses of the theorem. Then one may sum (3.6) over these, obtaining

$$a_{\mathbf{r}} \sim \sum_k \mathbf{z}^{(k)}(\bar{\mathbf{r}})^{-\mathbf{r}} \sum_{l \geq 0} b_{k,l}^*(\bar{\mathbf{r}}) |\mathbf{r}|^{-(l+d-1)/2}.$$

Theorem 3.19 is somewhat messy; even when $l = 1$, the combination of partial derivatives of G and H is cumbersome, though the prescription for each b_l in terms of partial derivatives of G and H is completely algorithmic for any l and d. In the applications in this paper, we will confine our asymptotic computations to the leading term. The following expression for the leading coefficient is more explicit than (3.6) and is proved in [57, Theorems 3.5 and 3.1].

THEOREM 3.20 (smooth point leading term). *When $G(\mathbf{z}(\bar{\mathbf{r}})) \neq 0$, the leading coefficient is*

$$b_0 = (2\pi)^{(1-d)/2} \mathcal{H}^{-1/2} \frac{G(\mathbf{z}(\bar{\mathbf{r}}))}{-z_d \partial H / \partial z_d}, \tag{3.7}$$

where $\mathcal{H}$ denotes the determinant of the Hessian of the function $\tilde{f}_{\mathbf{z}}$ at the origin.

COROLLARY 3.21. *In particular, when $d = 2$ and $G(\mathbf{z}(\overline{\mathbf{r}})) \neq 0$, we have*

$$a_{rs} \sim \frac{G(x,y)}{\sqrt{2\pi}} x^{-r} y^{-s} \sqrt{\frac{-yH_y}{sQ(x,y)}}, \tag{3.8}$$

where $(x,y) = \mathbf{z}(\overline{(r,s)})$ and $Q(x,y)$ is defined to be the expression

$$\begin{aligned} &-(xH_x)(yH_y)^2 - (yH_y)(xH_x)^2 \\ &- \left[(yH_y)^2 x^2 H_{xx} + (xH_x)^2 y^2 H_{yy} - 2(xH_x)(yH_y)xyH_{xy}\right]. \end{aligned} \tag{3.9}$$

□

Remark. In the combinatorial case, the expression in the radical will be positive real (this is true for more general F with correct choice of radical). The identity $ryH_y = sxH_x$ (see Proposition 3.11) shows that the given expression for a_{rs}, though at first sight asymmetric in x and y, has the expected symmetry.

3.4. Multiple Point Asymptotics. Throughout this section, $\mathbf{z}$ will denote a minimal multiple point. We will let m denote the number of sheets of $\mathcal{V}$ intersecting at $\mathbf{z}$ and let

$$H = \prod_{k=1}^{m} H_k^{n_k} \tag{3.10}$$

denote a local representation of H as a product of nonnegative integer powers of functions whose zero set is locally smooth. For $1 \leq k \leq m$, let $\mathbf{b}^{(k)}$ denote the vector whose jth component is $z_j \partial H_k / \partial z_j$.

We divide the asymptotic analysis of $a_\mathbf{r}$ into two cases, namely, $m \geq d$ and $m < d$. In the former case, since we have assumed transverse intersection at multiple points, the stratum of $\mathbf{z}$ is just the singleton $\{\mathbf{z}\}$. The asymptotics of $a_\mathbf{r}$ are simpler in this case. In fact it is shown in [56, Theorem 3.1] that

$$a_\mathbf{r} = \mathbf{z}^{-\mathbf{r}} \left(P(\mathbf{r}) + E(\mathbf{r})\right), \tag{3.11}$$

where P is piecewise polynomial and E decays exponentially on compact subcones of the interior of $\mathbf{K}(\mathbf{z})$. We begin with this case.

We will state three results in decreasing order of generality. The first is completely general, the second holds in the special case $m = d$ and $n_k = 1$ for all $k \leq m$, and the last holds for $m = d = 2$ and $n_1 = n_2 = 1$. Our version of the most general result provides a relatively simple formula from [4] but under the more general scope of [58].

DEFINITION 3.22. *Let $M = \sum_{k=1}^m n_k$ and, for $1 \leq j \leq M$, let t_j be integers so that the multiset $\{t_1, \ldots, t_M\}$ contains n_k occurrences of k for $1 \leq k \leq m$. Define a map $\Psi : \mathbb{R}^M \to \mathbb{R}^d$ by $\Psi(e_j) = \mathbf{b}^{(t_j)}$. Let λ^M be Lebesgue measure on $\mathcal{O}^M$ and let $P(\mathbf{x})$ be the density at $\mathbf{x}$ of the pushforward measure $\lambda^M \circ \Psi$ with respect to Lebesgue measure under Ψ. The function P is a piecewise polynomial of degree $M - d$, the regions of polynomiality being no finer than the common refinement of triangulations of the set $\{\mathbf{b}^{(k)} : 1 \leq k \leq m\}$ in $\mathbb{RP}^{d-1}$* [4, Definition 5].

The above definition of P is a little involved but is clarified by the worked example in section 4.11. Armed with this definition, we may state what happens when there are d or more sheets of $\mathcal{V}$ intersecting at a single point.

THEOREM 3.23 (isolated point asymptotics). *If $m \geq d$ and $\mathbf{z}$ is a minimal point in a singleton stratum with $G(\mathbf{z}) \neq 0$, then uniformly over compact subcones of $\mathbf{K}(\mathbf{z})$,*

$$a_\mathbf{r} \sim G(\mathbf{z}) P\left(\frac{r_1}{z_1}, \ldots, \frac{r_d}{z_d}\right) \mathbf{z}^{-\mathbf{r}}, \tag{3.12}$$

provided that $\mathsf{contrib}_{\overline{\mathbf{r}}}$ *contains only* $\mathbf{z}$*. This formula may be summed over* $\mathsf{contrib}_{\overline{\mathbf{r}}}$ *as long as* $\mathsf{contrib}_{\overline{\mathbf{r}}}$ *has finite cardinality.*

In the case $\mathbf{z} = \mathbf{1}$*,* (3.12) *reduces further to*

$$a_{\mathbf{r}} \sim G(\mathbf{1})P(\mathbf{r}) .$$

If, furthermore, $a_{\mathbf{r}}/G(\mathbf{1})$ *are integers, then* $a_{\mathbf{r}}/G(\mathbf{1})$ *is actually a piecewise polynomial whose leading term coincides with that of* $P(\mathbf{r})$*.*

Proof. First assume $\mathbf{z}$ is strictly minimal. Theorem 3.6 of [58] gives an asymptotic expression for $a_{\mathbf{r}}$ valid whenever $\mathbf{z}$ is a strictly minimal multiple point in a singleton stratum. The formula, while not impossible to use, is not as useful as the later formula given in [4, equation (3.8)]. This latter equation was derived under the assumption that H is a product of linear polynomials. Noting that the formula in [58, Theorem 3.6] depends on H only through the vectors $\{\mathbf{b}^{(k)} : 1 \le k \le m\}$, we see it must agree with [4, equation (3.8)], which is (3.12). The last statement of the theorem follows from (3.11).

To remove the assumption of strict minimality, suppose that $\mathbf{z}$ is minimal but not strictly minimal and that there are finitely many points $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(k)}$ in $\mathsf{contrib}_{\overline{\mathbf{r}}}$, necessarily all on a torus. The torus T in (3.2) may be pushed out to $(1+\epsilon)\mathbf{T}(\mathbf{z})$ except in neighborhoods of each $\mathbf{z}^{(j)}$. The same sequence of surgeries and residue computation in the proofs of [58, Proposition 4.1, Corollary 4.3, and Theorem 4.6] now establish Theorem 3.6 of [58] without the hypothesis of strict minimality, and as before one obtains (3.12) from equation (3.8) of [4]. □

While this gives quite a compact representation of the leading term, it may not be straightforward to compute P from its definition as a density. When $M = m = d$, P is a constant and the computation may be reduced to the following formula.

COROLLARY 3.24 (simple isolated point asymptotics). *If* $m = d$ *and each* $n_k = 1$ *in* (3.10)*, then if* $G(\mathbf{z}) \neq 0$ *and* $\mathbf{z}$ *is a minimal point in a singleton stratum, then*

$$a_{\mathbf{r}} \sim G(\mathbf{z})\mathbf{z}^{-\mathbf{r}} \left|J\right|^{-1} ,$$

where J *is the Jacobian matrix* $(\partial H_i/\partial z_j)$*. This formula may be summed over finitely many points as in the remark following Theorem* 3.19*.* □

Occasionally it is useful to have a result that does not depend on finding an explicit factorization of H. The matrix J can be recovered from the partial derivatives of H. The result of this in the case $m = d = 2$ is given by the following corollary [58, Theorem 3.1].

COROLLARY 3.25 (simple isolated point asymptotics, dimension 2). *If* $M = m = d = 2$*, and if* $G(\mathbf{z}) \neq 0$*, then setting* $\mathbf{z} = (x, y)$*,*

$$a_{rs} \sim x^{-r}y^{-s}\frac{G(x,y)}{\sqrt{-x^2y^2\mathcal{H}}},$$

where $\mathcal{H} := H_{xx}H_{yy} - H_{xy}^2$ *is the determinant of the Hessian of* H *at the point* (x, y)*. In the special but frequent case* $x = y = 1$*, we have simply*

$$a_{rs} \sim \frac{G(1,1)}{\sqrt{-\mathcal{H}}} .$$

This formula may be summed over finitely many points as in the second remark following Theorem 3.19*.* □

Finally, we turn to the case $m < d$. In this case, $\mathbf{z}$ is in a stratum containing more than just one point. The leading term of $a_{\mathbf{r}}$ is obtained by doing a saddle point integral of a formula such as (3.12) over a patch of the same dimension as the stratum. Stating the outcome takes two pages in [58]. Rather than give a formula for the resulting constant here, we state the asymptotic form and refer the reader to [58, Theorem 3.9] for evaluation of the constant, b_0.

THEOREM 3.26. *Suppose that as $\overline{\mathbf{r}}$ varies over a compact subset K of $\overline{\mathcal{O}}$, the set* $\mathsf{contrib}_{\overline{\mathbf{r}}}$ *is always a singleton varying over a fixed stratum of codimension $m < d$. If also $\mathbf{z}(\overline{\mathbf{r}})$ remains a strictly minimal multiple point, with each smooth sheet being a simple pole of F, and if $G(\mathbf{z}) \neq 0$ on K, then*

$$a_{\mathbf{r}} \sim \mathbf{z}(\overline{\mathbf{r}})^{-\mathbf{r}} b_0(\overline{\mathbf{r}}) |\mathbf{r}|^{\frac{m-d}{2}} \tag{3.13}$$

uniformly as $|\mathbf{r}| \to \infty$ in K. ∎

3.5. Distributional Limits. Of the small body of existing work on multivariate asymptotics, a good portion focuses on limit theorems. Consider, for example, the point of view taken in [5, 7] and the sequels to those papers [24, 8], where the numbers $a_{\mathbf{r}}$ define a sequence of $(d-1)$-dimensional arrays, the kth of which is the *horizontal slice* $\{a_{\mathbf{r}} : r_d = k\}$ (cf. section 1.2). Often this point of view is justified by the combinatorial application, in which the last coordinate, r_d, is a size parameter, and the goal is to understand rescaled limits of the horizontal slices. Distributional limit theory assumes nonnegative weights $a_{\mathbf{r}}$, so for the remainder of this section we assume we are in the combinatorial case, $a_{\mathbf{r}} \geq 0$.

If $C_k := \sum_{\mathbf{r}\,:\, r_d = k} a_{\mathbf{r}} < \infty$ for all k, we may define the slice distribution μ_k on $(d-1)$-vectors to be the probability measure giving mass $a_{\mathbf{r}}/C_k$ to the vector $(r_1, \ldots, r_{d-1})$ in the kth slice. There are several levels of limit theorem that may be of interest. A weak law of large numbers (WLLN) is said to hold if the measures $A \mapsto \mu_k(kA)$ converge to a point mass at some vector $\mathbf{m}$ (here, division of A by k means division of each element by k). Equivalently, a WLLN holds if and only if for some mean vector $\mathbf{m}$, for all $\epsilon > 0$,

$$\lim_{k \to \infty} \mu_k \left\{ \mathbf{r} : \left| \frac{\mathbf{r}}{k} - \mathbf{m} \right| > \epsilon \right\} = 0 \,. \tag{3.14}$$

Stronger than this is Gaussian limit behavior. As $\mathbf{r}$ varies over a neighborhood of size $\sqrt{k}$ about $k\mathbf{m}$ with r_d held equal to k, (3.6) takes the following form:

$$a_{\mathbf{r}} \sim C |\mathbf{r}|^{(-(d-1)/2} \exp\left[\phi(\mathbf{r})\right] \,.$$

Here the factor $C|\mathbf{r}|^{-(d-1)/2}$ varies only by $o(1)$ when $\mathbf{r}$ varies by $O(\sqrt{k})$, and

$$\phi(\mathbf{r}) := -\mathbf{r} \cdot \log |\mathbf{z}(\mathbf{r})| \,,$$

or simply $-\mathbf{r} \cdot \log \mathbf{z}(\mathbf{r})$ when $\mathbf{z}(\mathbf{r})$ is real. The WLLN identifies the location $\mathbf{r}_{\max}$ at which ϕ takes its maximum for r_d held constant at k. A Taylor expansion then gives

$$\exp\left[\phi(\mathbf{r}_{\max} + \mathbf{s}) - \phi(\mathbf{r}_{\max})\right] \sim \exp\left[-\frac{B(\mathbf{s})}{2k}\right] \tag{3.15}$$

for some nonnegative quadratic form, B. Typically B will be positive definite, resulting in a local central limit theorem for $\{\mu_k\}$.

Unfortunately, the computation of B, though straightforward from the first two Taylor terms of H, is very messy. Furthermore, it can happen that B is degenerate, and there does not seem to be a good test for this. As a result, existing limit theorems such as [7, Theorems 1 and 2] and [24, Theorem 2] all contain the hypothesis "if B is nonsingular." Because of the great attention that has been paid to Gaussian limit results, we will state a local central limit result at the end of this section. But since we cannot provide a better method for determining nonsingularity of B, we will not develop this in the examples.

Distributional limit theory requires the normalizing constants C_k to be finite. We make the slightly stronger assumption that $a_{\mathbf{r}} = 0$ when $\mathbf{r}/r_d$ is outside a compact set K. We define the map

$$\mathbf{r} \mapsto \mathbf{r}^* := \left(\frac{r_1}{r_d}, \ldots, \frac{r_{d-1}}{r_d}\right),$$

which will be more useful for studying horizontal slices than was the previously defined projection $\mathbf{r} \mapsto \hat{\mathbf{r}}$.

THEOREM 3.27 (WLLN). *Let F be a d-variate generating function with nonnegative coefficients and with $a_{\mathbf{s}} = 0$ for $\mathbf{s}/s_d$ outside some compact set K. Let $f(x) = F(1, \ldots, 1, x)$. Suppose f has a unique singularity x_0 of minimal modulus which is a simple pole. Let $\overline{\mathbf{r}} = \mathsf{dir}(\mathbf{x})$, where $\mathbf{x} = (1, \ldots, 1, x_0)$. If F is meromorphic in a neighborhood of $\mathbf{D}(\mathbf{x})$, then the measures $\{\mu_k\}$ satisfy a WLLN with mean vector $\mathbf{r}^* := (r_1/r_d, \ldots, r_{d-1}/r_d)$.*

Proof. We begin by showing that $\log\mathcal{D}$ has a unique normal at $\log|\mathbf{x}|$, which is in the direction $\overline{\mathbf{r}}$. We know that x_0 is positive real, since f has nonnegative coefficients. The proof of Theorem 3.16 showed that a neighborhood $\mathcal{N}$ of $\mathbf{x}$ in $\mathcal{V}$ maps, via log-moduli, onto a neighborhood of $\log|\mathbf{x}|$ in $\partial\log\mathcal{D}$. Since x_0 is a simple pole of f, we see that $\mathcal{N}$ is smooth with a unique normal $\nabla H(\mathbf{x})$, whence $\log\mathcal{D}$ has the unique normal direction $\overline{\mathbf{r}}$ at $\log|\mathbf{x}|$.

We observe next that $\sum_k C_k x^k = f(x)$, whence, via Theorem 2.2, $C_k \sim c x_0^{-k}$. Thus Theorem 3.27 will follow once we establish

$$\limsup_k \frac{1}{k} \log \left(\sum_{\mathbf{s}: s_d = k, \mathbf{s}^* \in K, |\mathbf{s}^* - \mathbf{r}^*| > \epsilon} a_{\mathbf{s}} \right) < -\log x_0 \,. \tag{3.16}$$

Each $\mathbf{s}^* \in K$ with $|\mathbf{s}^* - \mathbf{r}^*| \geq \epsilon$ is not normal to a support hyperplane at $\log|\mathbf{x}|$, so for each such $\mathbf{s}^*$ there is a $\mathbf{y} \in \partial\log\mathcal{D}$ for which

$$h_{\mathbf{s}^*}(\mathbf{y}) < h_{\mathbf{s}^*}(\mathbf{x}) = -\log x_0 \,. \tag{3.17}$$

By compactness of $\{\mathbf{s}^* \in K : |\mathbf{s}^* - \mathbf{r}^*| \geq \epsilon\}$ we may find finitely many $\mathbf{y}$ such that one of these, denoted $\mathbf{y}(\mathbf{s}^*)$, satisfies (3.17) for any $\mathbf{s}^*$. By compactness again,

$$\sup_{\mathbf{s}^*} h_{\mathbf{s}^*}(\mathbf{y}(\mathbf{s}^*)) < -\log x_0 \,.$$

It follows from the representation (3.2), choosing a torus just inside $\mathbf{T}(\mathbf{y})$ for $\mathbf{y}$ as chosen above depending on $\mathbf{s}$, that $|\mathbf{s}|^{-1} \log a_{\mathbf{s}}$ is at most $-\log x_0 - \delta$ for some $\delta > 0$ and all sufficiently large $\mathbf{s}$ with $|\mathbf{s}^* - \mathbf{r}^*| \geq \epsilon$. Summing over the polynomially many such $\mathbf{s}$ with $s_d = k$ proves (3.16) and the theorem. ∎

Coordinate slices are not the only natural sections on which to study limit theory. One might, for example, study limits of the sections $\{a_{\mathbf{r}} : |\mathbf{r}| = k\}$, that is, over a foliation of $(d-1)$-simplices of increasing size. The following version of the WLLN is adapted to this situation. The proof is exactly the same when slicing by $|\mathbf{r}|$ or any other linear function as it was for slicing by r_d.

THEOREM 3.28 (WLLN for simplices). *Let F be a d-variate generating function with nonnegative coefficients. Let $f(x) = F(x, \ldots, x)$. Suppose f has a unique singularity x_0 of minimal modulus which is a simple pole. Let $\overline{\mathbf{r}} = \mathsf{dir}(\mathbf{x})$, where $\mathbf{x} = (x_0, \ldots, x_0)$. If F is meromorphic in a neighborhood of $\mathbf{D}(\mathbf{x})$, then the measures $\{\mu'_k\}$, defined analogously to μ_k but over the simplices $\{|\mathbf{r}| = k\}$, satisfy a WLLN with mean vector $\hat{\mathbf{r}}$.* □

We end with a statement of a local central limit theorem (LCLT).

THEOREM 3.29 (LCLT). *Let $F, f, x_0, \mathbf{x}, \mathbf{r}$ and $\mathbf{r}^*$ be as in Theorem 3.27. Suppose further that $\mathsf{contrib}_{\overline{\mathbf{r}}}$ is the singleton, $\{\mathbf{x}\}$. If the quadratic form B from (3.15) is nonsingular, then*

$$\lim_{k\to\infty} k^{(d-1)/2} \sup_{\mathbf{r}} |\mu_k(\mathbf{r}) - \mathfrak{n}_B(\mathbf{r})| = 0, \tag{3.18}$$

where

$$\mathfrak{n}_B(\mathbf{r}) := (2\pi)^{-(d-1)/2} \det(B)^{-1/2} \exp\left(-\frac{1}{2k} B(\mathbf{r} - k\mathbf{r})\right) \tag{3.19}$$

is the discrete normal density. □

4. Detailed Examples. In sections 4–7 we work a multitude of examples. We aim to cover a sufficient variety of examples so that readers with a new application are likely to find a worked example that is similar to their own. With luck, the derivation of the result will be adaptable, and readers can thereby avoid original sources as well as the bulk of section 3 of the present survey. The use of computer algebra is essential to many of these examples, and we preface the worked examples with a brief explanation of this.

To apply any of the theorems of sections 3.3–3.5, one must solve for the point $\mathbf{z} \in \mathcal{V}$ as a function of the direction $\overline{\mathbf{r}}$ and then plug this into a variety of formulae. This may always be done numerically, but there are advantages to doing it analytically when possible: one may then differentiate, compute asymptotics near a point of interest, solve in terms of other data, and so forth. Even when one is interested chiefly in a numerical approximation for a single $a_{\mathbf{r}}$, one often obtains better numerics by simplifying analytically as much as possible.

When the generating function is rational or algebraic, as is quite frequent in applications, if one attempts to solve for $\mathbf{z}$ and plug in the results, it often turns out that a computer algebra system cannot simplify an expression that one suspects should be simpler. Consequently, the computer cannot tell whether such an expression is zero, may have trouble plotting the expression, and so forth. As is well known in computational commutative algebra, these problems can be avoided by working directly with polynomial ideals. This survey uses the Gröbner package in the computer algebra system Maple to illustrate the required computational algebra (accompanying Maple worksheets will be available in 2008 from http://www.math.upenn.edu/~pemantle/papers/papers.html/20.mws). We hope that these commands from Maple version 10 will remain in the Maple platform, but, in any case, we include platform-independent explanations of the algebra. Our treatment is necessarily quite brief, and readers who require more explanation should consult a source such as [15, 68].

Suppose a point $\mathbf{z} \in \mathbb{C}^d$ is the solution to polynomial equations $P_1(\mathbf{z}) = \cdots = P_k(\mathbf{z}) = 0$. The set of all polynomials P for which $P(\mathbf{z}) = 0$ is an *ideal* and, if the number of common solutions is finite, is said to be a *zero-dimensional* ideal. Such an ideal has many generating sets or *bases*, some of which are particularly useful to know. In particular, given a *term order*, that is, an order on the monomials $\mathbf{z}^{\mathbf{r}}$ that obeys certain properties, each ideal has a unique *Gröbner basis*. The lexicographic term order, known to Maple as `plex`, has the property that it forces the Gröbner basis of a zero-dimensional ideal to contain a univariate polynomial in the variable designated to come last in the lexicographical order. Thus, if p_1, p_2, and p_3 are polynomials in x, y, and z with finitely many common solutions, then the commands (the first just loads the Gröbner basis package into Maple)

```
> with(Groebner);
> L := Basis([p1,p2,p3],plex(x,z,y));
```

will produce a basis for the ideal of all polynomials vanishing on the common solutions to p_1, p_2, and p_3. The last variable in the variable list is y, so the choice of term order `plex` causes the last polynomial of the basis L to be a polynomial in y alone, whose roots are the possible y-coordinates of the common solution points.

A few words may prove helpful concerning the classification of critical points into smooth, multiple, and bad points. A stratum always consists of points of a single topological type. The critical point equations are different for different strata. Therefore, sorting critical points by type occurs as a preliminary step to sorting strata by type. If one assumes the denominator H of F is square free (if not, a single Maple command allows one to pass to the radical), then the smooth points of $\mathcal{V}$ are precisely those where ∇H does not vanish. The first step is always to check whether there are any nonsmooth points; generically, there are not, though interesting applications tend not to be generic. This step is accomplished by the single command

```
> Basis ([H , diff(H,x) , diff(H,y) , diff(H,z)] , tdeg(x,y,z));
```

Maple returns the trivial ideal `[1]` if and only if there are no nonsmooth points on $\mathcal{V}$. We have used `tdeg` rather than the slower `plex` to check whether we get `[1]`; if not, we can go back and use `plex` to produce a more useful basis for the ideal corresponding to the variety of singular points. Further tests may be done to determine whether a stratum of singular points consists of multiple points: one must check whether the radical of the ideal is *reduced*; for definitions and algorithms, the reader is referred to a text such as [53].

The algebra above is often sufficient to compute quantities such as $\mathsf{crit}_{\overline{\mathbf{r}}}$, which is defined by the polynomial equations (3.5). One may, however, go further: the substitution of one of these finitely many points into a polynomial expression such as (3.9) results in an algebraic number which Maple usually cannot write in its simplest form. Later, we discuss two reasonable ways to handle this. Given a quantity x defined by algebraic equations $p_1, \dots, p_k$, the most straightforward way to simplify the polynomial expression $Q(x)$ is to reduce Q modulo the ideal generated by $\{p_1, \dots, p_k\}$ (see section 4.2). An alternative is to obtain directly the minimal polynomial satisfied by $Q(x)$. This may be done by elimination (see section 4.4) or by matrix representation (see section 4.5). These techniques are useful for almost every example, but to avoid being repetitive, we often refer back to these fully worked computations.

4.1. Binomial Coefficients. To allow the reader a chance to check the application of Corollary 3.21 in a familiar setting, we begin with an example where the numbers

a_{rs} are explicitly known. Let $a_{rs} = \binom{r+s}{r,s} = \frac{(r+s)!}{r!\,s!}$ be the binomial coefficients and let

$$F(x,y) = \sum_{r,s\geq 0} a_{rs} x^r y^s \,.$$

The binomial coefficients satisfy the recurrence $a_{r,s} = a_{r-1,s} + a_{r,s-1}$, and this holds for all $(r,s) \neq (0,0)$, provided that we take a_{rs} to be zero when either r or s is negative. Linear recursions for a_{rs} in terms of values $a_{r',s'}$ with $(r',s') \leq (r,s)$ in the coordinatewise partial order lead easily to rational generating functions (as discussed in section 7, more general linear recursions do not yield rational functions). To find F, we observe from the recursion that all the coefficients of $(1-x-y)F$ vanish except the $(0,0)$-coefficient, which is 1. Thus

$$F(x,y) = \frac{1}{1-x-y} \,.$$

Let us compute the set $\mathsf{contrib}_{\overline{\mathbf{r}}}$ and the quantities appearing in Corollary 3.21. The singular variety $\mathcal{V}$ is the complex line $x+y=1$. The numerator of F is 1, and, in particular, it never vanishes. For any $\mathbf{z} \in \mathcal{V}$, the space $\mathbf{L}(\mathbf{z})$ is the linear span of $\mathbf{z}$. To see this, either use Proposition 3.11 or note that the tangent space to $\mathcal{V}$ is everywhere orthogonal to $(1,1)$ and plug this into Definition 3.10.

For each direction $\overline{\mathbf{r}}$ in the positive real orthant, there is thus a unique solution $\mathbf{z} \in \mathcal{V}$ to $\overline{\mathbf{r}} \in \mathbf{L}(\mathbf{z})$, namely, $\mathbf{z} = \hat{\mathbf{r}} = (\frac{r}{r+s}, \frac{s}{r+s})$ for any representative (r,s) of $\overline{\mathbf{r}}$. One may apply Theorem 3.16 to conclude that $\mathsf{contrib}_{\overline{\mathbf{r}}} = \{\hat{\mathbf{r}}\}$.

To apply Corollary 3.21, we need only compute the quantity Q in (3.9) and verify that it is nonzero. We have $H = 1-x-y, H_x = -1, H_y = -1$, and all other partial derivatives of H are zero, whence $Q = -xy(x+y)$; plugging in $(x,y) = (\frac{r}{r+s}, \frac{s}{r+s})$ gives

$$\frac{-yH_y}{sQ(x,y)} = \frac{1}{sx(x+y)} = \frac{r+s}{rs} \,.$$

Substituting this and $G \equiv 1$ into (3.8) gives

$$a_{rs} \sim \left(\frac{r+s}{r}\right)^r \left(\frac{r+s}{s}\right)^s \sqrt{\frac{r+s}{2\pi rs}} \,. \tag{4.1}$$

This asymptotic expression is valid as $(r,s) \to \infty$, uniformly if r/s and s/r remain bounded—see the uniformity conclusion in Theorem 3.19. In fact, we know from Stirling's formula that this holds uniformly as $\min\{r,s\} \to \infty$; see [49] for how to obtain the latter result in the present framework.

Although we have not presented precise error bounds in our asymptotic approximations, the errors for such smooth bivariate problems can be shown to be of the order $1/s$, and our approximations are good even for moderate values of r and s. As a randomly chosen example, we observe that the approximation above yields a relative error of about 0.8% when $r = 25, k = 12$, while the relative error has decreased to about 0.4% when $r = 50, k = 24$ and 0.08% when $r = 250, s = 120$.

4.2. Delannoy Numbers. Recall the Delannoy numbers a_{rs} that count paths from the origin to the point (r,s) with each step having displacement $(1,0)$, $(0,1)$, or $(1,1)$. In exactly the same way that we obtained the generating function for the binomial coefficients, we may use the recursion $a_{r,s} = a_{r-1,s} + a_{r,s-1} + a_{r-1,s-1}$ valid for all (r, s) except $(0, 0)$ to obtain the generating function

$$F(x,y) := \sum_{r,s\geq 0} a_{rs} x^r y^s = \frac{1}{1-x-y-xy},$$

so the denominator is given by $H = 1 - x - y - xy$.

Solving for $\mathsf{contrib}_{\overline{\mathbf{r}}}$ is only a little more involved for this generating function than it was for $1/(1-x-y)$. However, performing all calculations by hand, as we did in the case of binomial coefficients, is tedious for this example and completely impractical for later examples. The step-by-step computation below illustrates how a computer algebra system such as Maple can carry the derivation to completion at a symbolic level.

The linear space $\mathbf{L}(x, y)$ is the one-dimensional complex vector space spanned by $(xH_x, yH_y) = (-x(1 + y), -y(1 + x))$. The most convenient way in which to solve the equations $H = 0$ and $\overline{\mathbf{r}} \in \mathbf{L}(x, y)$ is to obtain a Gröbner basis. We load Maple's Gröbner package using the command `with(Groebner)` and then execute the following command:

```
> GB:=Basis([H, s*x*diff(H, x) - r*y*diff(H, y)], plex(x,y));
```

This computes a basis for the ideal corresponding to the common solutions to $H = 0$ and $\overline{(r, s)} \in \mathbf{L}(x, y)$. The answer is $GB := [-s + sy^2 + 2ry, s - sy - r + rx] =: [p_1(y), p_2(x, y)]$. The first element of the basis is an elimination polynomial in y alone. Solving $p_1 = 0$ yields the two values $y = (-r \pm \sqrt{r^2 + s^2})/s$. For each of the two y-values there is a unique x-value obtained by solving p_2 for x. One might observe that symmetry tells us these will be $(-s \pm \sqrt{r^2 + s^2})/r$; however, we need p_2 to tell us which x-value goes with which y-value. We find that there is a unique positive solution

$$x(\overline{\mathbf{r}}) = \frac{\sqrt{r^2+s^2}-s}{r};$$
$$y(\overline{\mathbf{r}}) = \frac{\sqrt{r^2+s^2}-r}{s}.$$

By Theorem 3.16 we see that $\mathsf{contrib}_{\overline{\mathbf{r}}} = (x(\overline{\mathbf{r}}), y(\overline{\mathbf{r}}))$ for any positive direction $\overline{\mathbf{r}}$. The numerator, G, is again identically 1. It remains only to plug the values for x and y into expressions for Q and a_{rs} and simplify.

Using the definition of Q and simplifying, we obtain $Q = xy(1+x)(1+y)(x+y)$. Let us reduce Q modulo H by the command

```
NormalForm (Q, [H, s*x*diff(H ,x) - r*y*diff(H, y)], plex(x,y));
```

to get

$$Q = -2x^2 - 2y^2 + 6x + 6y - 4\,. \tag{4.2}$$

Finally, substituting $x = x(\overline{\mathbf{r}})$ and $y = y(\overline{\mathbf{r}})$ into $-yH_y/(sQ)$ yields

$$\frac{-yH_y}{sQ} = \frac{rs}{\sqrt{r^2+s^2}(r+s-\sqrt{r^2+s^2})^2},$$

and putting this all together yields the expression (1.4) for the Delannoy numbers:

$$a_{rs} \sim \left(\frac{\sqrt{r^2+s^2}-s}{r}\right)^{-r} \left(\frac{\sqrt{r^2+s^2}-r}{s}\right)^{-s} \sqrt{\frac{1}{2\pi}} \sqrt{\frac{rs}{\sqrt{r^2+s^2}(r+s-\sqrt{r^2+s^2})^2}}\,.$$

For example, putting $r = s = n$ shows that the central Delannoy numbers a_{nn} have first order asymptotic approximation

$$a_{nn} \sim \frac{\cosh(\frac{1}{4}\log 2)}{\sqrt{\pi}}(3 + 2\sqrt{2})^n n^{-1/2}.$$

4.3. Powers, Quasi-powers, and Generalized Riordan Arrays. It often happens that we wish to estimate $[z^n]v(z)^k$, that is, the nth coefficient of a large power of a given function $v(z)$. Clearly, this is equal to the $x^n y^k$ coefficient of the generating function

$$F(x,y) := \frac{1}{1 - yv(x)}\,. \tag{4.3}$$

One place where this arises is in the enumeration of a combinatorial class whose objects are strings built from given blocks. Let $v(z) := \sum_{n=1}^{\infty} a_n z^n$ count the number a_n of blocks of size n. Then the generating function (4.3) counts objects of a given size by the number of blocks in the object.

EXAMPLE 4.1. *A long sequence of zeros and ones may be divided into blocks by repeatedly stripping off the unique initial string that is a leaf of T, a given* prefix tree. *Lempel–Ziv coding, for instance, does this but with an evolving prefix tree. When $v(x)$ is the generating function for the number a_n of leaves of T at depth n, then $1/(1 - yv(x))$ generates the numbers a_{rs} of strings of length r made of s blocks (the final block must be complete).*

Another place where generating functions of this form arise is in the Lagrange inversion formula. This application is discussed at length in section 6, but briefly, if h solves the equation $h(z) = zv(h(z))$, then even if we cannot explicitly solve for h, its coefficients are given by

$$[z^n]h(z) = \frac{1}{n}[z^{n-1}]v(z)^n\,.$$

This identity has been very profitable in the analysis of planar graphs and maps; cf. the discussion of results of [25, 3] in section 6.3.

A third place where coefficients of powers arise is in sums of independent random variables. Let $v(z) = \sum_{n=0}^{\infty} a_n z^n$ be the probability generating function for a distribution on the nonnegative integers, that is, $a_n = \mathbb{P}(X_j = n)$, where $\{X_j\}$ are a family of independent, identically distributed random variables. Then $v(z)^n$ is the probability generating function for the partial sum $S_n := \sum_{j=1}^{n} X_j$, and hence

$$\mathbb{P}(S_n = k) = [z^k]v(z)^n\,.$$

A *Riordan array* is defined to be an array $\{a_{nk} : n, k \geq 0\}$ whose generating function $F(x,y) := \sum_{n,k\geq 0} a_{nk} x^n y^k$ satisfies

$$F(x,y) = \frac{\phi(x)}{1 - yv(x)} \tag{4.4}$$

for some functions ϕ and v with $v(0) = 0$ and $\phi(0) \neq 0$. If in addition $v'(0) \neq 0$, the array is called a *proper* Riordan array. Just as (4.3) represents sums of independent, identically distributed random variables when v is a probability generating function, the format (4.4) generalizes this to *delayed renewal* sums (see, e.g., [19, section 3.4]), where an initial summand X_0 may be added that is distributed differently from the others. The quasi-powers (1.2) arising in GF-sequence analysis, which were described in section 1.2, are asymptotically of this form as well. Thus (4.4) approximately encompasses most of the known results leading to Gaussian behavior in multivariate generating functions.

Riordan arrays have been widely studied. In addition to enumerating a great number of combinatorial classes, Riordan arrays also behave in an interesting way under matrix multiplication (note that the condition $v(0) = 0$ implies $a_{nk} = 0$ for $k < n$, and, by triangularity of the infinite array, that multiplication in the Riordan group is well defined). Surveys of the Riordan group and its combinatorial applications may be found in [65, 64].

As we will see in this section, the asymptotic analysis of these arrays is relatively simple or, at least, is no more difficult than analyses of the functions ϕ and v that define the array. One should note, however, that Riordan arrays are often defined by data other than ϕ and v. Commonly, one has a linear recurrence for $a_{n,k+1}$ as a sum $\sum_{s=1}^{k-n} c_s a_{n+s,k}$; the generating function $A(t) := \sum_{j=1}^{\infty} c_j t^j$ is known and is fairly simple, but the function $v(x)$ in (4.4) is known only implicitly through the equation [63, equation (6)]

$$v(x) = xA(v(x)),$$

which is reminiscent of Lagrange inversion. The paper [72], which is devoted to bivariate asymptotics of Riordan arrays, discusses at length how to proceed when the known data includes $A(t)$ rather than $v(x)$. Thus there are versions of (4.10) below, such as Proposition 6.2, which state asymptotics in terms of A without explicit mention of v. The discussion in this section, however, will be limited to deriving asymptotics in terms of ϕ and v. For our analyses it is not important to require $v(0) = 0$ (for example, neither the binomial coefficient nor Delannoy number examples above satisfies that condition), so we drop this hypothesis and consider generalized Riordan arrays that satisfy (4.4) but may have $v(0) \neq 0$.

The following computations show that the quantities $\mathbf{L}(x, y)$ and $Q(x, y)$ turn out to be relatively simply expressed in terms of the function $v(x)$. Define the quantities

$$\mu(v; x) := \frac{xv'(x)}{v(x)}\,; \tag{4.5}$$

$$\sigma^2(v; x) := \frac{x^2 v''(x)}{v(x)} + \mu(v; x) - \mu(v; x)^2 = x\frac{d\mu(v; x)}{dx}\,. \tag{4.6}$$

It is readily established that for $(x, y) \in \mathcal{V}$, we have $\mathbf{L}(x, y) = \overline{(\mu(v; x), 1)}$. In other words, $(x, 1/v(x)) \in \mathsf{crit}_{(r,s)}$ if and only if $s\mu(v; x) = r$. Furthermore, when this holds, $Q(x, 1/v(x)) = \sigma^2(v; x)$. Provided that ϕ and σ^2 are nonzero at a minimal point, the leading term of its asymptotic contribution in (3.8) then becomes

$$a_{rs} \sim x^{-r} v(x)^s \frac{\phi(x, 1/v(x))}{\sqrt{2\pi s\sigma^2(v; x)}}, \tag{4.7}$$

where $\mu(v; x) = r/s$.

The notation μ and σ^2 is of course drawn from probability theory. These quantities are always nonnegative when v has nonnegative coefficients. To relate this to the limit theorems in section 3.5, observe that setting $x = 1$ gives

$$\mu(v; 1) = \frac{v'(1)}{v(1)}\,; \tag{4.8}$$

$$\sigma^2(v; 1) = \frac{vv'' - (v')^2 + vv'}{v^2}(1)\,. \tag{4.9}$$

Thus, under the hypotheses of Theorem 3.27, a WLLN will hold with mean $\mathbf{m} = \mu(v; 1)$. Of course we see here that $\mu(v; 1)$ is simply the mean of the renormalized distribution on the nonnegative integers with probability generating function v. Similarly, we see in Theorem 3.29 that $B(r, s) = (s - \mu(v; 1)r)^2/\sigma^2(v; 1)$ is the Gaussian term corresponding to the variance $\sigma^2(v; 1)$ of this renormalized distribution.

Suppose we are in the combinatorial, aperiodic case: the coefficients of v are nonnegative and $v(x)$ cannot be written as $x^b g(x^d)$ for any power series g and $d > 1$. The set $\mathcal{V}$ is the union of the set $\mathcal{V}_0$, parametrized as $(x, 1/v(x))$, with the union of horizontal lines where the value of x is a singular value for v or ϕ. As x increases from 0 to R, the (possibly infinite) minimum of the radii of convergence of v and ϕ, it is easy to verify that all the points of $\mathcal{V}_0$ encountered are strictly minimal and that $\sigma^2(v; x) > 0$. Since the derivative of $\mu(v; x)$ is $\sigma^2(v; x)/x$, this shows that $\mu(v; x)$ is strictly increasing on $(0, R)$. Thus $A := \mu(v; 0)$ and $B := \mu(v; R)$ are well defined as one-sided limits and for $A < \lambda < B$ there is a unique solution to $\mu(v; y) = \lambda$. In fact, A is the order of v at 0, and so equals 1 for proper Riordan arrays, but may be 0 for generalized Riordan arrays or an integer greater than 1 for improper Riordan arrays. From this it is evident that $a_{rs} = 0$ for $r/s < A$, so we should not have expected a solution to $\mu(v; x) = \lambda$ when $\lambda < A$. If $R = \infty$, then B is the (possibly infinite) degree of v, and one has again that $a_{rs} = 0$ for $r/s > B$. If v or ϕ is not entire, then one cannot say without further analysis what is expected for a_{rs} when $r/s \to \lambda > B$. We summarize in the following proposition.

PROPOSITION 4.2. *Let $(v(x), \phi(x))$ determine a generalized Riordan array. Suppose that $v(x)$ has radius of convergence $R > 0$ and is aperiodic with nonnegative coefficients, and that ϕ has radius of convergence at least R. Let A, B be as above. Then for $A < r/s < B$,*

$$a_{rs} \sim v(x)^s x^{-r} s^{-1/2} \sum_{k=0}^{\infty} b_k(r/s) s^{-k}, \tag{4.10}$$

where x is the unique positive real solution to $\mu(v; x) = r/s$. Here $b_0 = \phi(x)/\sqrt{2\pi\sigma^2(v; x)} \neq 0$. The asymptotic approximation is uniform as r/s varies within a compact subset of (A, B), whereas $a_{rs} = 0$ for $r/s < A$. □

We note that if the combinatorial restriction is lifted, much more complicated behavior can occur. The generating function $3/(3 - 3x - y + x^2)$ is of Riordan type with $\phi(x) = v(x) = (3 - 3x + x^2)^{-1}$, and even though v is aperiodic, $\mathsf{contrib}_{\mathbf{r}}$ has cardinality 2. Furthermore, at the unique contributing point for the diagonal direction, σ^2 vanishes.

The condition on the radius of convergence of ϕ is satisfied in most applications. One way in which it may fail is when F is a product of more than one factor. We will see an example of this in section 4.4. The next two subsections consider applications of Proposition 4.2 to combinatorial applications.

4.4. Maximum Number of Distinct Subsequences. Flaxman, Harrow, and Sorkin [22] consider strings over an alphabet of size d, which we take to be $\{1, 2, \ldots, d\}$

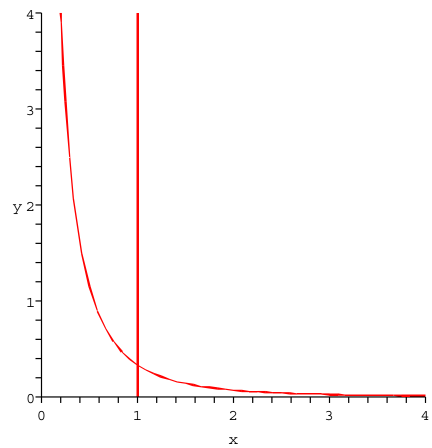

Fig. 4 $\mathcal{V}$ *in the case* $d = 3$.

for convenience. They are interested in strings of length n which contain as many distinct subsequences (not necessarily contiguous) of length k as possible. Let a_{nk} denote the maximum number of distinct subsequences of length k that can be found in a single string of length n. Initial segments $S|_n$ of the infinite string S consisting of repeated blocks of the string $12 \cdots d$ always turn out to be maximizers; that is, $S|_n$ has exactly a_{nk} distinct subsequences of length k. The generating function for $\{a_{nk}\}$ is given by [22, equation (7)]

$$F(x, y) = \sum_{n,k} a_{nk} x^n y^k = \frac{1}{1 - x - xy(1 - x^d)} .$$

This is of Riordan type with $\phi(x) = (1 - x)^{-1}$ and $v(x) = x + x^2 + \cdots + x^d$.

The case $d = 1$ is uninteresting. Suppose that $d \geq 2$. The singular variety $\mathcal{V}$ is the union of the line $x = 1$ and the smooth curve $y = 1/v(x)$, and they meet transversely at the double point $(1, 1/d)$; see Figure 4 for an illustration of this when $d = 3$.

This is a case where the radius of convergence of ϕ is less than the radius of convergence of v, the former being 1 and the latter being infinite. We have $\mu(v; x) = 1/(1 - x) - dx^d/(1 - x^d) = (1 + 2x + 3x^2 + \cdots + dx^{d-1})/(1 + x + x^2 + \cdots + x^{d-1})$. As x increases from 0 to 1 (the radius of convergence of ϕ, which is the value of x at the double point), μ increases from 1 to $(d + 1)/2$. Thus when $\lambda := n/k$ remains in a compact subinterval of $(1, \frac{d+1}{2})$, the Gaussian asymptotics of (4.10) hold.

To compute these in terms of λ, we solve for x in

$$\mu(v; x) = \lambda := \frac{n}{k} \tag{4.11}$$

and plug this into (4.7). One can do this numerically, but in the case where v is a polynomial, we can do better.

Solving $\mu(v; x) = \lambda$ by radicals and plugging into (4.6), which worked in section 4.2, will not be possible when $d \geq 5$ and is not practical even for smaller d. However, we see that σ^2 is algebraic, in the same degree $d - 1$ extension of the rationals that contains the value of x solving (4.11). We look therefore for a polynomial with coefficients in $\mathbb{Q}(\lambda)$, of degree $d-1$, which annihilates the σ^2 in Proposition 4.2. To find this polynomial, the best tactic is to work directly with generators of polynomial ideals, and we give the details below.

When $d = 2$, the solution

$$x(\lambda) = \frac{\lambda - 1}{2 - \lambda}$$

is a rational function of λ and nothing fancy is needed to arrive at $\sigma^2 = (\lambda-1)(2-\lambda)$. We therefore illustrate with $d = 3$, though this procedure is completely general and will work any time v is a polynomial.

Plugging the expression (4.5) for $\mu(v;x)$ into (4.11) and clearing denominators gives a polynomial equation for x,

$$x\frac{dv}{dx} - \lambda v = 0\,.$$

In our example,

$$x\left(1 + 2\,x + 3\,x^2\right) - \lambda\,\left(x + x^2 + x^3\right) = 0\,. \tag{4.12}$$

We now need to evaluate

$$\sigma^2(v;x) = x\frac{d\mu}{dx} = \frac{x\left(1 + 4\,x + x^2\right)}{\left(1 + x + x^2\right)^2} \tag{4.13}$$

at the value x that solves (4.12). To do this we compute a Gröbner basis of the ideal in $\mathbb{Q}(\lambda)[x,S]$ generated by $\mu(v;x) - \lambda$ and $\sigma^2(v;x) - S$ (after clearing denominators). The commands

```
p1:=(1+2*x+3*x^2)-lambda*(1+x+x^2):
p2:=x*(1+4*x+x^2)-S*(1+x+x^2)^2:
Basis([p1, p2], plex(x, S));
```

produce the elimination polynomial

$$p(S;\lambda) = 3S^2 + (6\lambda^2 - 24\lambda + 16)S + 3\lambda^4 - 24\lambda^3 + 65\lambda^2 - 68\lambda + 24,$$

which is easily checked to be irreducible (using Maple's `factor` command, for example), and hence is generically the minimal polynomial for σ^2. It is easy to choose the right branch of the curve: the function $x(\lambda)$ increases in $(0,1)$ as λ increases in $(1,2)$, and σ^2 is given in (4.13) as an explicit function of x that is easily checked to be increasing. It follows from these results that σ^2 increases from 0 to 2/3 as λ goes from 1 to 2.

To finish describing the asymptotics, we first note that values of λ greater than d are uninteresting. It is obvious that any prefix of S of length at least dk will allow all possible k-subsequences to occur. Thus $a_{nk} = d^k$ when $\lambda \geq d$.

We already know that as $\lambda := n/k \to (d+1)/2$ from below, the asymptotics are Gaussian, and the exponential growth rate approaches d. For slopes $\lambda \geq (d+1)/2$, we use Theorem 3.16 to see that for each such λ, there is a minimal point in the positive quadrant controlling asymptotics in direction λ. The only minimal point of $\mathcal{V}$ we have not yet used is the double point $(1, 1/d)$. It is readily computed that this cone has extreme rays corresponding to $\lambda = (d+1)/2$ and $\lambda = \infty$, and thus asymptotics in the interior of the cone will be supplied by the double point. Using Corollary 3.25 we obtain $a_{\lambda k,k} \sim d^k$.

Although section 3 did not discuss what happens when λ approaches but is not equal to $(d+1)/2$, some results are known. A refinement of Corollary 3.25, given in [58, Theorem 3.1 (ii)], shows that asymptotics in the boundary direction $\lambda = (d+1)/2$ are smaller by a factor of 2 than in the interior of the cone. In fact an examination of the proof there (see [58, Lemma 4.7 (ii)]) shows that one has Gaussian behavior,

$$a_{nk} \sim d^k \Phi(x) \text{ for } n = \frac{d+1}{2}k + x\sqrt{\frac{d^2-1}{12}\,k},$$

where $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,dt$ is the standard normal cumulative distribution function (CDF) and the constant $\sqrt{(d^2-1)/12}$ is obtained in a manner similar to Q. This can also be obtained from a probabilistic analysis as follows. The quantity $d^{-k}a_{nk}$ is the probability that a uniformly chosen sequence of length k is a subsequence of $S|_n$. The length of an initial substring of S required to contain a given sequence $u_1u_2\cdots u_k$ is $\sum_{j=1}^{k}(u_j - u_{j-1})$ mod d, where 0 mod d is taken to be d and where $u_0 := 0$. Thus the probability of a uniformly chosen word of length k being a subsequence of $S|_n$ is equal to

$$\mathbb{P}\left(\sum_{j=1}^{k} U_j \le n\right),$$

where U_j are independent uniform picks from $\{1,\ldots,d\}$. The central limit theorem now gives

$$\mathbb{P}\left(\sum_{j=1}^{k} U_j \le \frac{d+1}{2}k + x\sqrt{\frac{d^2-1}{12}\,k}\right) \sim \Phi(x)\,.$$

4.5. Paths, Hills, and Fine Numbers. *Dyck paths* are paths from the origin whose steps are in the set $\{(1,1),(1,0),(1,-1)\}$ and that never go below the x-axis. In section 7.3 we will count Dyck paths by their final point, which may be anywhere in the bottom half of the first quadrant. Often in the literature the term Dyck path is reserved for a path constrained to end on the x-axis. In [17], a number of combinatorial interpretations are found for these constrained Dyck paths when counted by final x-value and by the number of *hills*, a hill being a peak of height 1 (a peak is an occurrence of the step $(1,1)$ immediately followed by $(1,-1)$). Let a_{nk} denote the number of Dyck paths from the origin to $(2n,0)$ that have k hills. Of particular interest are the values a_{n0} which count hill-free Dyck paths and are called *Fine numbers*.

The generating function for $\{a_{nk}\}$ is derived in [17, Proposition 4]:

$$F(x,y) := \sum_{n,k} a_{nk}x^n y^k = \frac{2}{1+2x+\sqrt{1-4x}-2xy} = \frac{v(x)/x}{1-yv(x)}\,, \tag{4.14}$$

with

$$v(x) = \frac{1-\sqrt{1-4x}}{3-\sqrt{1-4x}} = \frac{2x}{1+\sqrt{1-4x}+2x}\,.$$

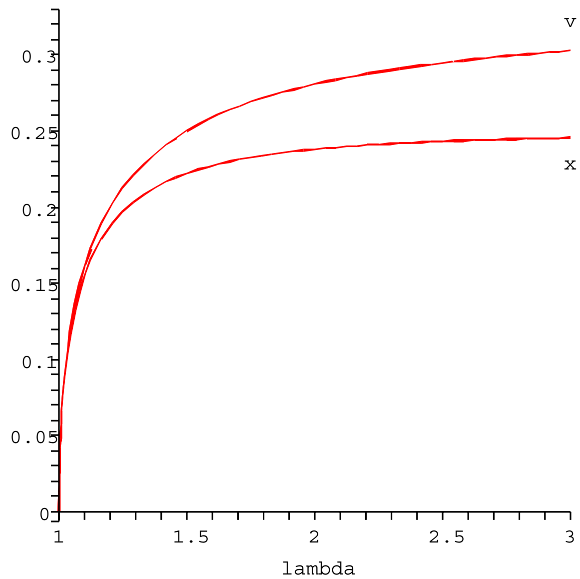

Fig. 5 *x and v plotted against λ.*

Rationalizing the denominator, we have

$$v(x) = \frac{1 + 2x - \sqrt{1 - 4x}}{4 + 2x}\,, \tag{4.15}$$

and thus the domain of convergence is given by $|x| < 1/4, |yv(x)| < 1$. For $1 < \lambda < \infty$ the critical points are determined by

$$\mu(v; x) = \frac{4x}{(1 - \sqrt{1 - 4x})(\sqrt{1 - 4x})(3 - \sqrt{1 - 4x})} = \lambda\,.$$

The left side of the above equality increases strictly from 1 at $x = 0$ to ∞ as $x \uparrow 1/4$, and there is a unique positive real solution x_λ for each $\lambda > 1$. Figure 5 shows x increasing from 0 to 1/4 and v increasing from 0 to 1/3 as λ increases from 1 to ∞.

We now want to complete the computations by simplifying $\sigma^2(v; x)$ at the value of x satisfying $\mu(v; x) = \lambda$. We sketch here how the computations of section 4.4 generalize to the algebraic case. The algebra is only a little more involved than it is for rational v, but some changes are clearly required.

In order to perform Gröbner basis computations, we must first pass from (4.15) to the implicit form $\alpha(x, v) = 0$, where α is a polynomial in $\mathbb{Q}[x, v]$. It is easy in this case to see by inspection that

$$\alpha(v, x) = (2 + x)v^2 - (1 + 2x)v + x\,;$$

at worst, implicitizing will require going back to the original derivation and following the computation at the level of ideals.[2] Next, according to (4.5) and (4.6), differenti-

[2]Implicitization techniques for polynomial or rational parametrizations $x = f(t), v = g(t)$, though not needed here, may be found in [15, sections 3.2 and 6.4].

ating α implicitly gives μ and σ^2 as rational functions of v and x:

$$
\begin{aligned}
v' &= -\frac{\alpha_x}{\alpha_v} = \frac{v^2 - 4x}{2v(2+x)}, \\
v'' &= -\frac{\alpha_v^2\alpha_{xx} + \alpha_x^2\alpha_{vv} - 2\alpha_{xv}\alpha_x\alpha_v}{\alpha_v^3} = \frac{3v^4 + 16v^2 - 16x^2}{4v^3(2+x)^2}, \\
\mu &= -\frac{x\alpha_x}{v\alpha_v} = \frac{x(4x - v^2)}{2v^2(2+x)}, \\
\sigma^2 &= -\frac{x\left(xv\alpha_v^2\alpha_{xx} + xv\alpha_x^2\alpha_{vv} - 2xv\alpha_{xv}\alpha_x\alpha_v + v\alpha_v^2\alpha_x + x\alpha_x^2\alpha_v\right)}{\alpha_v^3 v^2} \\
&= \frac{-x(8x^3 - 4x^2v^2 - 8xv^2 + v^4)}{v^4(2+x)^2}.
\end{aligned}
\tag{4.16}
$$

Setting $\mu = n/k$ and clearing denominators gives a polynomial $\beta = nv\alpha_v + kx\alpha_x \in \mathbb{Q}[n,k]\,[x,v]$ that vanishes when $\mu(v;x) = n/k$. In our present example, we have

$$\beta(x,v) = (4\,n + 2\,nx + kx)\,v^2 + (-2\,kx - n - 2\,nx)\,v + kx\,.$$

To find the critical point corresponding to the direction n/k, we could simply solve the equations $\alpha = 0, \beta = 0$ for x, v. Then to simplify σ^2, we add in the equation stating that $\sigma^2 = S$ (with denominator cleared) and find a Gröbner basis using the "plex" order with S as the last variable, just as in the previous section. Unfortunately, as is well known, such computations can be very slow, and in the present instance we have difficulty obtaining an answer with Maple in a reasonable amount of time.

Thus we use the following alternative method, as described in [15, section 2.2]. The polynomials α and β have finitely many common solutions (x,v) for any fixed n,k, so they define a zero-dimensional ideal, J. In other words, there are finitely many linearly independent monomials in x and v over $\mathbb{Q}[n,k]\,/\,\langle\alpha,\beta\rangle$. A convenient choice is the set of monomials not divisible by any leading term of a fixed Gröbner basis for J. With respect to this basis, one may express x and v as matrices over $\mathbb{Q}[n,k]$ for multiplication operators, then compute the matrix for σ^2 as a rational function of these matrices, and, finally, find the minimal polynomial γ for this matrix. The polynomial $\gamma \in \mathbb{Q}[n,k]$ will vanish at σ^2.

To carry this out in Maple, we use the following commands:

```
sys:=[alpha, beta]:
monomialorder := tdeg(x, v):
gb := Basis(sys, monomialorder):
ns, rv := NormalSet(gb, monomialorder):
Mx := MultiplicationMatrix(y, ns, rv, gb, monomialorder):
Mv := MultiplicationMatrix(v, ns, rv, gb, monomialorder):
```

Now evaluate the rational expression for σ^2 given above, with x, v replaced by M_x, M_v (making liberal use of the simplification capabilities of Maple). We then compute the minimal polynomial using the `MinimalPolynomial` command. The result is

$$
\begin{aligned}
&z^3\left(32n^4k^5\right) \\
&+z^2\left(-144n^7k^2 + 160n^6k^3 - 16n^5k^4 - 6n^4k^5 + 4n^2k^7 + 2k^9\right) \\
&+z\left(144n^9 - 304n^8k + 203n^7k^2 - 46n^6k^3 - 15n^5k^4 + 20n^4k^5 - 11k^6n^3 + 10n^2k^7 - nk^8\right) \\
&+\ \left(-15n^7k^2 + 31n^5k^4 - 27n^9 - 35n^6k^3 - 21n^4k^5 + 57n^8k + 11k^6n^3 - n^2k^7\right),
\end{aligned}
$$

and computing eigenvalues (using `solve`) we obtain three possible values for σ^2, namely,

$$S_1 := \frac{(n-k)(n+k)(3n^2+k^2)}{16n^4},$$

$$S_2 := \frac{\left(9n-k+3\sqrt{9n^2-10nk+k^2}\right)(n-k)\,n}{4k^3},$$

$$S_3 := \frac{\left(9n-k-3\sqrt{9n^2-10nk+k^2}\right)(n-k)\,n}{4k^3}.$$

Calculating which of these three gives the correct value for σ^2 may be done similarly to how it was done in the previous section. Equation (4.16) gives σ^2 as a univalent function of x and v. The correct branches for x and v as functions of $\lambda = n/k$ are shown in Figure 5. Plugging these into (4.16) and comparing to S_1, S_2, and S_3 (to avoid a numerical comparison, one may compare limits at 0) shows that the second expression is correct:

$$\sigma^2 = S_3 := \frac{\left(9n-k+3\sqrt{9n^2-10nk+k^2}\right)(n-k)\,n}{4k^3}.$$

To write $x(\lambda)$ and $v(\lambda)$ in their simplest forms, observe that the values of x and v are simply eigenvalues of the multiplication matrices M_x, M_v and so we obtain their respective minimal polynomials:

$$4n^2x^2 + (7n-k)(n-k)x - 2n(n-k)\,,$$
$$2kv^2 + 3(n-k)v - (n-k)\,.$$

Since we know that the relevant point has positive coordinates, we have the explicit forms

$$x = \frac{(3n-k)\sqrt{(n-k)(9n-k)} - (7n-k)(n-k)}{8n^2},$$

$$v = \frac{\sqrt{(n-k)(9n-k)} - 3(n-k)}{4k}.$$

As can be seen from this example, the computations can be rather messy, though still routine and automatable. In the end we have verified that $\sigma^2 > 0$ for $\lambda > 1$, and we conclude that the Gaussian asymptotics given by Proposition 4.2 hold uniformly as n/k varies over any compact subinterval of $(1, \infty)$.

Another way to simplify computations such as this one will be discussed in section 6. Of course, for any given value of n, k, we may shortcut the above computation and solve numerically for x to obtain σ^2. For example, with $n/k = 2$ we have $x = (5\sqrt{17}-13)/32 \approx 0.2379852541$, so that $y = 1/v(x) = (3+\sqrt{17})/2 \approx 3.561552813$. Hence

$$a_{2k,k} \sim (0.1228255460\cdots)(4.957474791\cdots)^k k^{-1/2}.$$

As far as we know, asymptotics such as these have not previously been computed. Using 10 significant figure floating point approximations, we obtain (using Maple) an answer accurate to within 0.8% for $k = 30$.

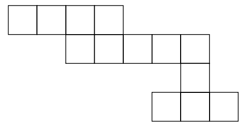

Fig. 6 *An HCP with* 13 *cells and* 4 *rows.*

4.6. Horizontally Convex Polyominoes. A *horizontally convex polyomino* (HCP) is a union of cells $[a, a+1] \times [b, b+1]$ in the two-dimensional integer lattice such that the interior of the figure is connected and every row is connected (see Figure 6). Formally, if $S \subseteq \mathbb{Z}^2$ and $P = \bigcup_{(a,b) \in S} [a, a+1] \times [b, b+1]$, then P is an HCP if and only if the following three conditions hold: $B := \{b : \exists a, (a, b) \in S\}$ is an interval; the set $A_b := \{a : (a, b) \in S\}$ is an interval for each $b \in \mathbb{Z}$; and whenever $b, b+1 \in B$, the sets A_b and A_{b+1} intersect. Let a_n be the number of HCPs with n cells, counting two as the same if they are translates of one another. Pólya [60] proved that

$$\sum_n a_n x^n = \frac{x(1-x)^3}{1 - 5x + 7x^2 - 4x^3} \,. \tag{4.17}$$

Further discussion of the origins of this formula and its accompanying recursion may be found in [55] and [66]. The proof in [71, pp. 150–153] shows in fact that

$$F(x, y) = \sum_{n,k} a_{nk} x^n y^k = \frac{xy(1-x)^3}{(1-x)^4 - xy(1 - x - x^2 + x^3 + x^2 y)} \,, \tag{4.18}$$

where a_{nk} is the number of HCPs with n cells and k rows. Let us find an asymptotic formula for a_{rs}.

All the coefficients of $F(x, y)$ are nonnegative; they vanish when $s > r$ but otherwise are at least 1. By Corollary 3.18 (the last part, which requires only $a_{\mathbf{r}} \geq 0$), we know that all points of $\mathcal{V}$ in the first quadrant that are on the southwest facing part of the graph (that is, that are minimal in the coordinatewise partial order) are contributing critical points. We will see later that there are no other critical points on each torus. As $\bar{\mathbf{r}}$ varies over $\mathbf{\Xi} = \{\bar{\mathbf{r}} : 0 < s/r < 1\}$ from the horizontal to the diagonal, the point $\mathsf{contrib}_{\bar{\mathbf{r}}}$ moves along this graph from $(1, 0)$ to $(0, \infty)$ (see Figure 7).

To make the mapping from $\bar{\mathbf{r}}$ to $\mathbf{z}$ explicit, we use the fact that $\bar{\mathbf{r}} \in \mathbf{K}(\mathbf{z}) \subseteq \mathbf{L}(\mathbf{z})$ (Theorem 3.16 and part (i) of Proposition 3.14), so $\mathbf{z}$ may be obtained from $\mathbf{r}$ by (3.5). It is readily computed that $\nabla H \neq \mathbf{0}$ except at $(1, 0)$. Thus all minimal points are smooth. The only solution to $G = H = 0$ is at $(1, 0)$, so the numerator is nonvanishing at any minimal point as well. Checking whether the quantity Q defined in (3.9) of Theorem 3.21 ever vanishes, we find that the only solutions to $H = Q = 0$ are at $(1, 0)$ and at complex locations that are not minimal because $\mathbf{L}(\mathbf{z})$ is not real there (Proposition 3.12). Last, we must check that there are no contributing points other than the minimal points in the positive quadrant. We can ascertain that this is true by checking the "extraneous" critical points, which will lie on three other branches of a quartic (see below), and seeing that they lie on the wrong torus. We would then know that the asymptotics for a_{rs} are uniform as s/r varies over a compact subset of the interval $(0, 1)$ and given by

$$a_{rs} \sim C x^{-r} y^{-s} r^{-1/2} \,.$$

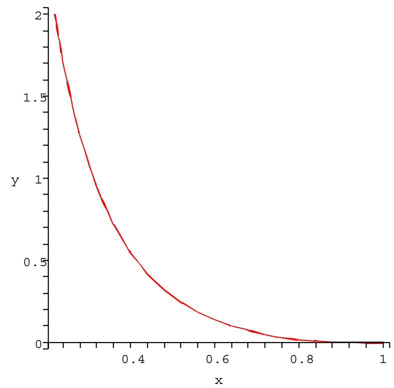

Fig. 7 *Minimal points of $\mathcal{V}$ in the positive real quadrant.*

We will use Maple to determine x, y, and C as explicit functions of $\lambda := s/r$, giving asymptotics for the number of HCPs whose shape is not asymptotically vertical or horizontal, but first we see what we can deduce without much computation.

A crude approximation at the logarithmic level is

$$a_{rs} \approx \exp\left[r(-\log x - (s/r)\log y\right],$$

where x and y of course still depend on s/r. We first compute the average row length of a typical HCP. We may apply Theorem 3.27 to find the limit in probability of h_k/k, where h_k is the height of an HCP chosen uniformly from among all HCPs of size k. Setting $y = 1$ in the bivariate generating function recovers the univariate generating function (4.17). The point $(x, 1)$, where $x = x_0$ is the smallest root of the denominator of (4.17), controls asymptotics in this direction; we compute $\mathsf{dir}(x, 1)$ there to be $(xH_x, yH_y)(x_0, 1)$, which simplifies to $r/s = 4(5 - 14x_0 + 12x_0^2)/(5 - 9x_0 + 11x_0^2)$ or still further to $\alpha := \frac{1}{47}(147 - 246x_0 + 344x_0^2) \approx 2.207$. We conclude that for large k, $k/h_k \to \alpha$. Thus we see that the average row length in a typical large HCP is around 2.2. Finally, let us see how the computer algebra for general $\overline{\mathbf{r}}$ turns out. To find (x, y) given $\overline{\mathbf{r}}$, solve the pair of equations $H = 0, sxH_x = ryH_y$; this find points on $\mathcal{V}$ with $\overline{\mathbf{r}} = \overline{(r, s)} \in \mathbf{L}(x, y)$. Explicitly, we set $\lambda = s/r$ and ask Maple for a Gröbner basis for the ideal generated by H and $\lambda xH_x - yH_y$. The following Maple code fragment is useful in such cases:

```
Hx := diff(H,x): Hy := diff(H,y): X:=x*Hx: Y:=y*Hy:
Hxx := diff(Hx,x): Hxy := diff(Hx,y): Hyy := diff(Hy,y):
Q := -X^2*Y-X*Y^2-x^2*Y^2*Hxx-X^2*y^2*Hyy+2*X*Y*x*y*Hxy:
L := [H,lambda*X-Y]:
gb := Basis(L, plex(y,x)):
```

Maple returns a basis consisting of polynomials α and $(x-1)^5\beta$, where the quartic

$$\beta := (1 + \lambda)x^4 + 4(1 + \lambda)^2x^3 + 10(\lambda^2 + \lambda - 1)x^2 + 4(2k - 1)^2x + (1 - \lambda)(1 - 2\lambda)$$

is the elimination polynomial for x and is generically irreducible. Furthermore α is linear in y. Rather than express y in terms of x, it is perhaps easier to compute the elimination polynomial for y, which is y^3 times the following polynomial:

$$(4\lambda^4 - 4\lambda^3 - 3\lambda^2 + 4\lambda - 1)y^4 + (40\lambda^4 - 44\lambda^3 - 20\lambda^2 + 48\lambda - 16)y^3 \\ + (-172\lambda^4 + 128\lambda^3 + 160\lambda^2 - 256\lambda + 64)y^2 + (1152\lambda^4 - 1024\lambda^3 - 512\lambda^2)y - 1024\lambda^4 .$$

Note that when $H = 0$, $x = 1$ if and only if $y = 0$. The point $(1, 0)$ is a solution to the critical point equations for every value of λ. However, it is never a contributing point, because $h_{\hat{\mathbf{r}}} = \infty$ when any coordinate vanishes (recall $\hat{\mathbf{r}}$ is in a compact subset of the positive orthant).

Thus generically we have 4 candidates for contributing points. Precisely one of these is minimal and in the first quadrant. The others do not contribute for generic $\overline{\mathbf{r}}$, which may easily be checked for any given $\overline{\mathbf{r}}$; a verification for all $\overline{\mathbf{r}}$ simultaneously would require more computer algebra.

Finally, from Corollary 3.21 we see that

$$C = \frac{xy(1-x)^3}{\sqrt{2\pi}} \sqrt{\frac{y\left(-x\left(1 - x - x^2 + x^3 + x^2 y\right) - x^3 y\right)}{Q}} .$$

The minimal polynomial for $\sqrt{2\pi}C$ can be computed as in previous sections; it turns out to have degree 8 for generic λ. Of course, given floating point approximations for x and y, we may simply compute an approximation for C directly.

As an example, suppose that $n = 2k$ so that $\lambda = 1/2$. In this case we have simplification and the minimal polynomials for x and y, respectively, are $3x^2 + 18x - 5$ and $75y^2 - 288y + 256$. Note that there is a single element

$$(x_0, y_0) := \left(\sqrt{\frac{32}{3}} - 3, \frac{48 - \sqrt{512}}{25}\right) \approx (0.265986, 1.397442)$$

of crit in the positive quadrant. By Theorem 3.16 this belongs to contrib, so by Proposition 3.8, contrib is the singleton $\{(x_0, y_0)\}$. We obtain (by naive floating point computation in Maple) $a_{n,n/2} \sim (0.237305\ldots)(3.18034\ldots)^n n^{-1/2}$. For $n = 60$, the relative error in this first order approximation is about 1.5%. Contrast the exponential growth rate with the exponential growth rate 3.20557 for the number of HCPs with n cells and any number of rows.

4.7. Symmetric Eulerian Numbers. The symmetric Eulerian numbers $\hat{A}(r, s)$ [14, p. 246] count the number of permutations of the set $[r+s+1] := \{1, 2, \ldots, r+s+1\}$ with precisely r descents. By reading backwards, this equals the number with exactly s descents, hence the symmetry in r and s. The symmetric Eulerian numbers have exponential generating function [32, section 2.4.21]

$$F(x, y) = \frac{e^x - e^y}{xe^y - ye^x} = \sum_{r,s} \frac{\hat{A}(r, s)}{r!\, s!} =: \sum_{r,s} a_{rs} x^r y^s . \tag{4.19}$$

The denominator in this representation is singular at the origin and has a factor of $(x - y)$ in common with the numerator. We factor this out of the numerator and denominator, writing $F = G/H$ with $G = (e^x - e^y)/(x - y)$ and $H = (xe^y - ye^x)/(x - y)$. In this representation, both the numerator and the denominator are entire functions. We quickly check that in the positive real quadrant, $\mathcal{V}$ is the graph of a monotone decreasing function as in the left-hand side of Figure 8 and that the quantity Q (see the Maple code fragment in section 4.6) does not vanish on $\mathcal{V}$:

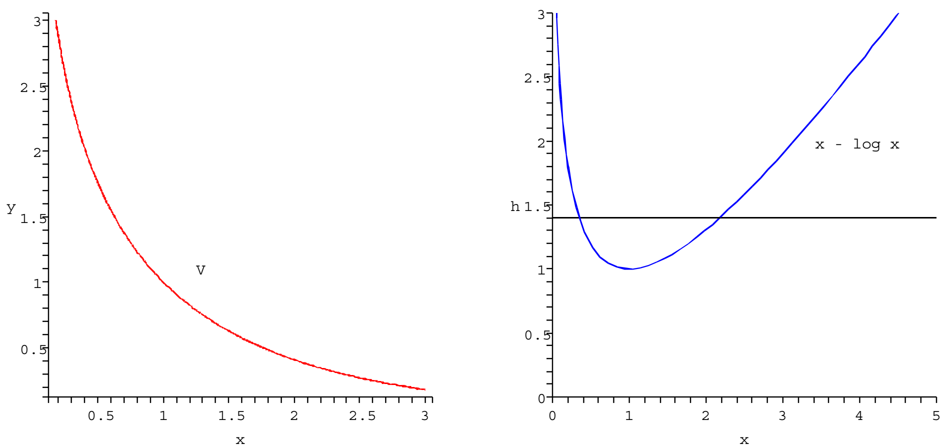

Fig. 8 *The positive real points of $\mathcal{V}$ and their description by height-pairs.*

- first check that F is not entire but is analytic in a neighborhood of the origin;
- next, l'Hôpital's rule shows that H never vanishes on $\{x = y\}$ except at $(1, 1)$;
- rearranging terms then shows that H vanishes when x and y are two positive real numbers with the same height $h(t) := t - \log t$ (see the right-hand side of Figure 8);
- the exact symbolic expression for Q has limit $e^3/12$ at $(1, 1)$ and this is the minimum value of Q on $\mathcal{V}$.

The symmetric Eulerian numbers are nonnegative, so we can find minimal points of $\mathcal{V}$ in the first quadrant. It is easy to check that the gradient of H never vanishes on $\mathcal{V}$, so $\mathcal{V}$ is smooth. At the point $x = y = 1$, we have $\partial H/\partial x = \partial H/\partial y = 1$, so $\mathsf{dir}(1, 1) = \mathbf{1}$. For any other point of $\mathcal{V}$, we may work with the representation of H having $(x - y)$ in the denominator. We set

$$\alpha = \frac{x\partial H/\partial x}{x\partial H/\partial x + y\partial H/\partial y}$$

so that $(x, y) \in \overline{\mathbf{r}} \Leftrightarrow \hat{\mathbf{r}} = (\alpha, 1 - \alpha)$. The expression for α does not look all that neat, but on $\mathcal{V}$ we may substitute $(y/x)e^x$ for e^y, after which the expression for α reduces to $(1 - x)/(y - x)$. We see therefore from Theorem 3.19 that $\hat{A}(r, s)$ is asymptotic to $C_\alpha(r + s)^{-1/2} r!s!\gamma^{r+s}$, where for a given value of $\alpha = r/(r + s)$, the value of γ is given by $x^{-\alpha}y^{-(1-\alpha)}$ after solving the following transcendental equations for (x, y):

$$\begin{aligned} \frac{1 - x}{y - x} &= \alpha\,; \\ xe^y &= ye^x\,. \end{aligned} \tag{4.20}$$

We do not know whether there is a closed form expression for γ in terms of α, though we note that by using (4.20) several times, one may simplify it somewhat to $\gamma = (y/x)^\alpha y^{-1} = \exp(\alpha(y - x))/y = \exp(1 - x)/y$.

The fact that the equations (4.20) have a unique positive real solution is possible to verify directly, but also follows from nonnegativity of the coefficients and the fact that the positive real part of $\mathcal{V}$ lies along the boundary of the domain of convergence. One must still check that $\mathbf{T}(x(\alpha), y(\alpha))$ contains no other points of $\mathcal{V}$. This is true for generic $\bar{\mathbf{r}}$ and can easily be checked for a given $\bar{\mathbf{r}}$, but a general proof that it works for all $\bar{\mathbf{r}}$ is not obvious.

4.8. Smirnov Words. Given an integer $d \geq 3$ we define a Smirnov word in the alphabet $\{1, \ldots, d\}$ to be a word in which no letter repeats consecutively. The number of Smirnov words of length n is of course easily seen by a direct counting argument to be $d(d-1)^{n-1}$. If we count these words according to the number of occurrences of each symbol, we get the multivariate generating function

$$F(\mathbf{z}) = \sum a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}} = \frac{1}{1 - \sum_{j=1}^{d} \frac{z_j}{1+z_j}}, \tag{4.21}$$

where $a_{\mathbf{r}}$ is the number of Smirnov words with r_j occurrences of the symbol j for all $1 \leq j \leq d$. This generating function may be derived from the generating function

$$E(\mathbf{y}) = \frac{1}{1 - \sum_{j=1}^{d} y_j}$$

for all words as follows. Note that collapsing all consecutive occurrences of each symbol in an arbitrary word yields a Smirnov word; this may be inverted by expanding each symbol of a Smirnov word into an arbitrary positive number of identical symbols, whence $F(\frac{y_1}{1-y_1}, \ldots, \frac{y_d}{1-y_d}) = E(\mathbf{y})$. Substituting $y_j = z_j/(1+z_j)$ yields the formula (4.21) for F; see [21, section II.7.3, Example 23] for more on this old result.

When we write F as a rational function, we see that

$$G = \prod_{j=1}^{d}(1+z_j) = \sum_{j=0}^{d} e_j(\mathbf{z}),$$
$$H = \sum_{j=0}^{d}(1-j)e_j(\mathbf{z}),$$

where e_j is the jth elementary symmetric function of $z_1, \ldots, z_d$ and $e_0 = 1$ by definition. Thus when $d = 3$, for example, one obtains $H = 1 - (xy + yz + xz) - 2xyz$. The expression for H may be denoted quite compactly by

$$H = g - u \frac{\partial g}{\partial u}\bigg|_{u=1},$$

where $g(\mathbf{z}, u) := \prod_{j=1}^{d}(1 + uz_j)$ is the polynomial in u whose coefficients are the elementary symmetric functions in $\mathbf{z}$. When $d \geq 3$, the denominator is of the form $1 - P$ for an aperiodic polynomial P with nonnegative coefficients, so by Proposition 3.17, $\mathsf{contrib}_{\bar{\mathbf{r}}}$ will always consist of one strictly minimal point in the positive orthant.

The typical statistics of a Smirnov word of length n are not in doubt, since it is clear that each letter will appear with frequency $1/d$. This may be formally deduced from Theorem 3.28, which also shows the variation around the mean to be Gaussian. We may, perhaps, be more interested in the so-called large-deviation probabilities: the exponential rate at which the number of words decreases if we alter the statistics. Let $\hat{\mathbf{s}}$ denote some frequency vector. Then, as $|\mathbf{r}| \to \infty$ with $\hat{\mathbf{r}} \to \hat{\mathbf{s}}$, we have

$$\frac{1}{|\mathbf{r}|} \log a_{\mathbf{r}} \to -\mathbf{r} \cdot \log |\mathbf{z}|,$$

where $\mathbf{z} \in \mathcal{O}^d$ satisfies $\mathsf{dir}(\mathbf{z}) = \bar{\mathbf{r}}$.

C	C	A	G	T	C	A	G	C	T	A
C		A	G	T	C		C	C	T	G
G	T	A	G	T	C			T	T	C
1	1	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	0	0	1	1	1

Fig. 9 *A* $(3; 11, 9, 9)$*-alignment and its corresponding binary matrix.*

To solve for $\mathbf{z}$ in terms of $\bar{\mathbf{r}}$, we use the symmetries of the problem. It is evident that z_j is symmetric in the variables $\{r_i : i \neq j\}$. For example, when $d = 3$, if we refer to $\mathbf{z}$ as (x, y, z) and $\hat{\mathbf{r}}$ as (r, s, t), then the equations (3.5) have the solution

$$x = \frac{r^2 - (s-t)^2}{2r(s+t-r)},$$

where the values of y and z are given by the same equation with r, s, and t permuted. When $d \geq 4$, the solution is not a rational function and it is difficult to get Maple to halt on a Gröbner basis computation. This points to the need for computational algebraic tools better suited to working with symmetric functions.

4.9. Alignments of Sequences. The problem of *sequence alignment* is of interest in molecular biology [70, 61], since a given string may evolve via substitutions, insertions, and deletions. We seek to place several strings of varying lengths in parallel, which may necessitate adding some spaces. A basic problem underlying the design of algorithms which seek the best alignment in the sense of minimizing some score function is simply to count such alignments. Note that for this problem the elements of the string are irrelevant—we only care whether there is a letter or a space.

Mathematically, given positive integers k and a k-tuple $\mathbf{n} = (n_1, \ldots, n_k)$, we may define a $(k; \mathbf{n})$-alignment as a $k \times n$ binary matrix for some n, such that no columns are identically zero and the ith row sum is n_i. Column j is *aligned* if it has no zero entry, and a *block* of size b is a $k \times b$ submatrix, with contiguous columns, all of which are aligned. For example, Figure 9 shows an alignment of 3 strings, and the corresponding matrix. The generating function giving the number of $(k, \mathbf{n})$-alignments with the ith row having sum n_i is given by

$$F(z_1, \ldots, z_k) = \frac{1}{2 - p(\mathbf{z})},$$

where $p(\mathbf{z}) = \prod_{j=1}^{k} (1 + z_j)$.

The special case where all sequences involved have the same length (all rows have the same sum) is easily dealt with. By Proposition 3.17, there is a single strictly minimal point in the positive orthant that controls asymptotics in the diagonal direction. By symmetry of F and of the equation $\mathsf{dir}(\mathbf{z}) = \mathbf{1}$, this point has the form $\mathbf{z} = z\mathbf{1}$ for some positive z. Thus we have $z = 2^{1/k} - 1$ and hence the asymptotic has the form $C(z^{-k})^n n^{-\frac{k-1}{2}}$. This result was derived by Griggs, Hanlon, Odlyzko, and Waterman in [33] using a saddle point analysis. Note that for large k, $z^{-k} \sim 2^{-1/2} k^k (\log 2)^{-k}$.

To compute the constant C, we use the formula of Theorem 3.20. It is readily computed that $-z_k \partial H / \partial z_k = 2z/(1+z)$. To compute the Hessian determinant, we need only consider the Hessian in $\theta_1, \dots, \theta_{d-1}$ of $-\log\prod_{j<k}(1+z\exp(i\theta_j)) + \log(2 - \prod_{j<k}(1+z\exp(i\theta_j))$. Direct computation using the fact that $\prod_{j<k}(1+z\exp(i\theta_j))$ takes the value $2/(1+z)$ when $\theta = \mathbf{0}$ shows that the Hessian matrix has diagonal elements $2/(1+z)$ and off-diagonal elements $1/(1+z)$. Thus $\mathcal{H} = k(1+z)^{-(k-1)} = k2^{1/k}/2$ and this yields

$$C = \frac{2^{(1-k^2)/2k}}{(2^{1/k}-1)\sqrt{k\pi^{k-1}}}\,.$$

This yields the result of [33] (note that there is an error in the formula displayed on [70, p. 1989]—the factor r^k should read r and the factor $2^{(k^2-1)/(2k)}$ should be $2^{(1-k^2)/(2k)}$).

A more important case for biological applications is when the minimum block size is bounded below, by b, say. The generating function in this case is shown in [61], by means of standard operations on generating functions of formal languages, to have the form

$$F(z_1,\dots,z_k) = \frac{A}{1+(1-p)A+(A-1)t} = \frac{1-t+t^b}{(1-t)(1-g)-t^b g} = \frac{1+\frac{t^b}{1-t}}{1-g\left(1+\frac{t^b}{1-t}\right)},$$

where $t = \prod_j z_i$, $g = p-1-t$, and $A = 1-t+t^b$. Note that when $b=1$, then $A=1$, and we recover the unrestricted case analyzed above. In this case, asymptotics for the case where all sequences have equal length have been derived (as far as we know) only for the case $k=2$, using the "diagonal method" discussed in section 8. However, we can readily deal with general k using the methods of this paper.

Again, by symmetry we need only look for a contributing minimal point of the form $\mathbf{z} = z\mathbf{1}$. Let $H = 1+(1-p)A+(A-1)t$. Since $t = z^k$ and $p = (1+z)^k$, we must find the root(s) ρ of smallest modulus of $h(z) := H(z,z,\dots,z)$. Note that the third formula above shows that we may take $h = 1-P$, where P is aperiodic with nonnegative coefficients, so there will be a unique root ρ of smallest modulus, which will be positive real. The exponential growth rate is then ρ^{-k} with polynomial correction of order $n^{(1-k)/2}$ as above.

We note that in the case $k=2$, such a result was proved in [34], but stated slightly differently. The value $\tau = \rho^2$ is given as the minimal positive root of the polynomial $(1-x)^2 - 4x(1-x+x^b)^2$. Setting $x = z^2$ and factoring the difference of squares yields ρ as the minimal positive root of $H(z,z) = 1-2z-z^2+2z^3-2z^{2b+1}$.

To better estimate $\rho := \rho_b$, note that $\rho_b < 1$ so it is reasonable to consider the approximation obtained by setting all powers involving z^b to zero. The minimal real zero ρ_1 of $(1-t)(1-g)$ is asymptotically of order $1/k$ (and equals $2^{1/k}-1$). Now $\rho_b > \rho_1$, but ρ_b should be close to ρ_1. Indeed, it appears that an iteration scheme based on the fixed point equation

$$z = \left(1+z^k+\frac{1-z^k}{1-z^k+z^{kb}}\right)^{1/k} - 1$$

given by $h(z) = 0$ converges rapidly to ρ_b from starting point ρ_1.

4.10. Probability That There Is an Edge in an Induced Subgraph. From the n-set $[n] := \{1,\dots,n\}$, a collection of t disjoint pairs is named. Then a k-element subset, $S \subseteq [n]$, is chosen uniformly at random. What is the probability $p(n,k,t)$ that

S fails to contain as a subset any of the t pairs? This question is posed in [51] as a step in computing the diameter of a random Cayley graph of a group of cardinality n when k elements are chosen at random (the diameter is infinite if the k elements do not generate G).

There are a number of ways of evaluating $p(n, k, t)$, one of which is by inclusion-exclusion on the number of pairs contained. This leads to

$$p(n,k,t) = \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{k-2i} \binom{n}{k}^{-1} = \sum_{i=0}^{t} (-1)^i \binom{t}{i} \binom{n-2i}{n-k} \binom{n}{k}^{-1}. \tag{4.22}$$

Here, we define the binomial coefficient $\binom{p}{q}$ to be zero unless $0 \le q \le p$.

The numbers $a(n,k,t) := \binom{n}{k} p(n,k,t)$ are simpler to analyze via generating functions; since for most purposes only a first order asymptotic is sought, we lose nothing in considering $a(n,k,t)$. Note that from the description above, $a(n,k,t) = 0$ if $k+t > n$ by the pigeonhole principle (since the complement of S has size less than t, S must contain at least $t+1$ of the $2t$ chosen elements). From the statement of the problem, $a(n,k,t)$ is not defined if $2t > n$; however, formula (4.22) still makes sense in that case, even though it does not define the probability of any event. In fact, $a(n,k,t)$ can be negative for large t.

The most direct approach to finding a trivariate generating function of $a(n,k,t)$ is to use some well-known bivariate generating functions $\sum_{i,j} a_{ij} x^i y^j$. If $a_{ij} = \binom{i+j}{j}$, then the generating function is $(1-x-y)^{-1}$, while that for $a_{ij} = \binom{i}{j}$ is $(1-x(1+y))^{-1}$. We now compute

$$\begin{aligned}
\sum_{n,k,t,i} x^n y^k z^t w^i \binom{t}{i} \binom{n-2i}{k-2i} &= \sum_{N,K,i,j} x^{N+2i} y^{K+2i} z^{i+j} w^i \binom{i+j}{i} \binom{N}{K} \\
&= \left(\sum_{i,j} \binom{i+j}{i} (zwx^2y^2)^i z^j \right) \left(\sum_{N,K} \binom{N}{K} x^N y^K \right) \\
&= \frac{1}{1-z(1+wx^2y^2)} \frac{1}{1-x(1+y)},
\end{aligned}$$

which yields the trivariate generating function

$$F(x,y,z) = \sum_{n,k,t} a(n,k,t) x^n y^k z^t = \frac{1}{1-z(1-x^2y^2)} \frac{1}{1-x(1+y)}. \tag{4.23}$$

Note that if we impose the restriction $2t \le n$, then the sum over N, K is restricted to $N \ge 2j$. Now summing over N, K, t, i as above we obtain the restricted trivariate generating function

$$F_2(x,y,z) = \sum_{\{n,k,t\colon\, 2t \le n\}} a(n,k,t) x^n y^k z^t = \frac{1}{1-x(1+y)} \frac{1}{1-zx^2(1+2y)}.$$

The advantage of F_2 is that all its coefficients are nonnegative.

We show how to give asymptotics for $a(n,k,t)$ in an arbitrary direction for which $n > k+t$. For the specific case of $k = \lfloor cn \rfloor, t = (n-4)/12$ with $0 < c < 1$, such an analysis was posed as an open question at the end of an early draft of [51].

It turns out that the computations can be carried out equally well with F or F_2. However, with F_2 we can use Theorem 3.16 if we are interested only in the behavior of $a(n,k,t)$ for $2t \le n$.

The denominator of F factors into two smooth pieces, call them $H_1 := 1-x(1+y)$ and $H_2 := 1 - z(1-x^2y^2)$. There is a corresponding stratification of $\mathcal{V}$ into two surfaces, $\mathcal{V}_1 := \{H_1 = 0 \neq H_2\}$ and $\mathcal{V}_2 = \{H_2 = 0 \neq H_1\}$, and a curve, $\mathcal{V}_0 := \{H_1 = H_2 = 0\}$. For $\mathbf{z} \in \mathcal{V}_1$, lack of dependence on z means that $\mathsf{dir}(\mathbf{z}) \perp (0,0,1)$, so dir cannot be in the strictly positive orthant; hence for $\overline{\mathbf{r}} \in \overline{\mathcal{O}}$, there are no points of $\mathsf{crit}_{\overline{\mathbf{r}}}$ in $\mathcal{V}_1$. It turns out there are no points of $\mathsf{crit}_{\overline{\mathbf{r}}}$ in $\mathcal{V}_2$ either. This is discovered by computation. If $\mathbf{z} \in \mathcal{V}_2 \cap \mathsf{crit}_{\overline{\mathbf{r}}}$ and $\overline{\mathbf{r}} = \overline{(r,s,1)}$, then $\mathbf{z}$ satisfies the equations

$$H_2(\mathbf{z}) = 0\,;$$
$$x\frac{\partial H_2}{\partial x} - rz\frac{\partial H_2}{\partial z} = 0\,;$$
$$y\frac{\partial H_2}{\partial y} - sz\frac{\partial H_2}{\partial z} = 0\,.$$

These equations turn out to have no solutions: Maple tells us in an instant that the ideal generated by the left-hand sides of the three equations above in $\mathbb{C}[x,y,z,r,s]$ contains $r - s$; thus $\mathcal{V}_2$ may contain points of $\mathsf{crit}_{\overline{\mathbf{r}}}$ only when $r = s$, that is, only governing asymptotics of $a(n,k,t)$ for which $n = k$, which are not interesting.

Evidently, $\mathsf{crit}_{\overline{\mathbf{r}}} \subseteq \mathcal{V}_0$. Two equations are $H_1 = H_2 = 0$, that is,

$$(x,y,z) = \left(\frac{1}{1+y}\,,\,y\,,\,\frac{(1+y)^2}{1+2y}\right). \tag{4.24}$$

This last equation says that $\overline{\mathbf{r}}$ is in the linear space spanned by $\nabla_{\log} H_1$ and $\nabla_{\log} H_2$. Setting the determinant of $(\overline{\mathbf{r}}, \nabla_{\log} H_1, \nabla_{\log} H_2)$ equal to zero gives the equation

$$2(r-s-1)y^2 + (r-3s)y - s = 0. \tag{4.25}$$

Note that exactly the same equations are obtained for x, y, z when using F_2.

If $r > s+1$, the discriminant $(r+s)^2 - 8s$ of the quadratic in (4.25) is positive and so (4.25) has one positive and one negative root, the positive root being

$$y_+ := \frac{\sqrt{(r+s)^2 - 8s} - (r-3s)}{4(r-s+1)}\,.$$

Plugging this into (4.24) for x_+ and z_+ yields expressions for these which are also quadratic over $\mathbb{Z}[r,s]$. The expression for (x_-, y_-, z_-) is just the algebraic conjugate of (x_+, y_+, z_+), but is negative in the second and third coordinates.

Having identified $\mathsf{crit}_{\overline{\mathbf{r}}}$ as these two conjugate points, which we will call $\mathbf{z}_+$ and $\mathbf{z}_-$, it remains to find $\mathsf{contrib}_{\overline{\mathbf{r}}}$. Since the two elements lie on different tori, we may conclude from Theorem 3.16 that $\mathsf{contrib}_{\overline{\mathbf{r}}} = \{\mathbf{z}_+\}$.

The form of the leading term asymptotic is then given by Theorem 3.26:

$$a(n,k,t) \sim C\left(\frac{k}{n}, \frac{t}{n}\right) n^{-1/2} x_+^{-n} y_+^{-k} z_+^{-t}\,.$$

For example, in the direction $\overline{(12, 12c, 1)}$ with $c < 11/12$, we can compute the exponential growth rate and compare with that of $\binom{n}{\lfloor cn \rfloor}$, and thereby show that certain random Cayley graphs have diameter 2 with very high probability. For details, see [51].

Analysis in the unrestricted case where we may have $2t > n$ is more difficult. In fact $[x^n y^k z^t]F$ need not be zero when $k + t > n$ and $2t > n$, and may be negative.

4.11. Integer Solutions to Linear Equations. Let $a_{\mathbf{r}}$ be the number of nonnegative integer solutions to $A\mathbf{x} = \mathbf{r}$, where A is a $d \times m$ integer matrix. Denote by $\mathbf{b}^{(k)}$ the kth column of A. Then

$$F(\mathbf{z}) = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}} = \prod_{k=1}^{m} \frac{1}{1 - \mathbf{z}^{\mathbf{b}^{(k)}}}.$$

This enumeration problem has a long history. We learned of it from [16] (see also [66, section 4.6]). One special case is when $A = A_{m,n}$ is the incidence matrix for a complete bipartite graph between m vertices and n vertices. Solutions to $A\mathbf{x} = \mathbf{r}$ count nonnegative integer $m \times n$ matrices with row and column sums prescribed by $\mathbf{r}$; enumerating these is important when constructing statistical tests for contingency tables. Another special case is when $A = A_n$ is the incidence matrix of a complete directed graph on n vertices, directed via a linear order on the vertices. Here solutions are counted in various cones other than the positive orthant, and the function enumerating them is known as Kostant's partition function.

It is known that $a_{\mathbf{r}}$ is piecewise polynomial, and a benchmark problem in computation has been to determine these polynomials, and the regions or *chambers* of polynomiality, explicitly. Indeed, several subproblems merit benchmark status. Counting the chambers for Kostant's partition function is one such problem. Another is to evaluate the leading term for the diagonal polynomial $a_{n\mathbf{1}}$ in the case where $A = A_{k,k}$, the so-called *Ehrhart polynomial*, which counts $k \times k$ nonnegative integer matrices with all rows and columns summing to n. Equivalently, this counts integer points in the n-fold dilation of the *Birkhoff polytope*, defined as the set in $\mathbb{R}^{\binom{k}{2}}$ of all doubly stochastic $k \times k$ matrices; the leading term of $a_{n\mathbf{1}}$ is the volume of the Birkhoff polytope.

Our methods do not improve on the computational efficiency of previous researchers: our endpoint is a well-known representation, from which other researchers have attempted to find efficient means of computing. Our methods do, however, give an effective answer, which illustrates that this class of problems can be put in the framework for which our methods give an automatic solution. The remainder of this section is devoted to such an analysis.

Our examination of the pole variety $\mathcal{V}$ begins with an answer rather than a question: we see that $\mathbf{1} \in \mathcal{V}$ and realize that life would be easy if $\mathsf{contrib}_{\overline{\mathbf{r}}} = \{\mathbf{1}\}$ for all $\overline{\mathbf{r}}$. In order for $\mathbf{1}$ to be a singleton stratum it is necessary and sufficient that the columns $\mathbf{b}^{(k)}$ span all of $\mathbb{R}^d$. For the remainder of the section we assume this to be the case. This satisfies our standing assumption 3.6.

The variety $\mathcal{V}$ is the union of the smooth sheets $\mathcal{V}_k$, where for $1 \leq k \leq m$, $\mathcal{V}_k$ is the binomial variety $\{\mathbf{z} : H_k(\mathbf{z}) := \mathbf{z}^{\mathbf{b}^{(k)}} - 1 = 0\}$. On each of these varieties, $\mathsf{dir}_k(\mathbf{z}) = \nabla_{\log}(H_k)(\mathbf{z})$ is constant and equal to $\mathbf{b}^{(k)}$. If a stratum S is bigger than a singleton, then at any $\mathbf{z} \in S$, the vectors $\nabla_{\log} H_k$ span a proper subspace of $\mathbb{R}^d$; but these vectors do not vary as $\mathbf{z}$ varies in S, so the union over $\mathbf{z} \in S$ of $\mathbf{L}(\mathbf{z})$ is this same proper subspace of $\mathbb{R}^d$. Consequently, the union over all points of all nonsingleton strata of $\mathbf{L}(\mathbf{z})$ is a union of proper subspaces, which we denote $\mathbf{\Xi}'$. For $\overline{\mathbf{r}} \notin \mathbf{\Xi}'$, then, $\mathsf{contrib}_{\overline{\mathbf{r}}}$ consists of one or more singleton strata.

Taking logs, we see that $\log \mathcal{V}_k$ is a hyperplane normal to $\mathbf{b}^{(k)}$ and is central (passes through the origin). We see that $\mathbf{K}(\mathbf{0})$ is the positive hull of the vectors $\mathbf{b}^{(k)}$; this hull is $\mathbf{\Xi}$ and outside its closure, $a_{\mathbf{r}}$ vanishes. Forgetting about the logs, we see that $\mathbf{1} \in \mathsf{contrib}_{\overline{\mathbf{r}}}$ for all $\overline{\mathbf{r}} \in \mathbf{\Xi}_0 := \mathbf{\Xi} \setminus \mathbf{\Xi}'$; this follows from Theorem 3.23 since $\overline{\mathbf{r}} \in \mathbf{K}(\mathbf{1})$.

If the intersection of all the surfaces $\mathcal{V}_k$ contains any other points on the unit torus $\mathbf{T}(\mathbf{1})$, then these too are in $\mathsf{contrib}_{\overline{\mathbf{r}}}$ for all $\overline{\mathbf{r}} \in \mathbf{\Xi}_0$. This is easy to check, since it is equivalent to the integer combinations of the columns of A spanning a proper sublattice of $\mathbb{Z}^d$. For instance, in the example of counting matrices with constrained row and column sums, the columns of A span the alternating sublattice of $\mathbb{Z}^d$, corresponding to the fact that $\mathsf{contrib}_{\bar{\mathbf{r}}} = \{\mathbf{1}, -\mathbf{1}\}$ for all $\bar{\mathbf{r}} \in \Xi_0$.

There may be singleton strata given by intersections of subfamilies of $\{\mathcal{V}_k : 1 \leq k \leq m\}$ that lie on the unit torus, but in computing the leading term asymptotics these may be ignored because they yield polynomials in $\mathbf{r}$ of lower degree.

In summary, for any $\bar{\mathbf{r}} \in \Xi_0$, the leading term asymptotics are given by summing (3.12) over a set containing $\mathbf{1}$ and isomorphic to the quotient of $\mathbb{Z}^d$ by the integer span of the columns of A. Our a priori knowledge that $a_{\mathbf{r}}$ are integers, teamed with (3.11) as in the last part of Theorem 3.23, shows that in fact $a_{\mathbf{r}}$ are piecewise polynomial, at least away from Ξ'. Thus, except on a set of codimension 1, we recover the well-known piecewise polynomiality of $a_{\mathbf{r}}$.

In their paper [16], de Loera and Sturmfels use as a running example the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} .$$

In this example we see that the columns of A span $\mathbb{Z}^3$ over $\mathbb{Z}$, so the only contributing point is $\mathbf{1}$. The first three $\mathbf{b}$ vectors, in the order given, are the standard basis vectors, so the cone Ξ is the whole positive orthant of $\mathbb{RP}^2$, which is a 2-simplex. The other two $\mathbf{b}$ vectors are two of the three face diagonals of this cone, which are two midpoints of edges of the 2-simplex in $\mathbb{RP}^2$. In addition to the boundary of Ξ, there are three projective line segments in Ξ', corresponding to the one-dimensional strata $\mathcal{V}_4 \cap \mathcal{V}_5$, $\mathcal{V}_4 \cap \mathcal{V}_3$, and $\mathcal{V}_5 \cap \mathcal{V}_2$. The complement of these three line segments (two medians of the simplex $\mathbb{RP}_2$ and the line segment connecting two midpoints of edges) divides $\mathbb{RP}^2$ into five chambers, and on each of these chambers $a_{\mathbf{r}}$ is a quadratic polynomial. Algorithms for computing these polynomials are given, for example, in [16] or [4, section 5].

4.12. Queuing Theory. Queuing theory describes the evolution of a collection of *jobs* that enter the system, exit the system, and change types stochastically as they are worked on by one or more *servers*. There are many variations on this model. The example we consider here is a closed network with no class hopping, meaning that no jobs enter, leave, or change type. The model is defined in terms of the parameters $J, K, \{\lambda_j : 1 \leq j \leq J\}, \{\mu_{jk} : 1 \leq j \leq J, 1 \leq k \leq K\}$, and $\{p_{j;kl} : 1 \leq j \leq J, 0 \leq k, l \leq K\}$ as follows.

Let J be the number of types of jobs. Let $K+1$ be the number of servers. In our example, server k $(1 \leq k \leq K)$ distributes its resources evenly among all jobs in its queue, serving jobs of type j at rate μ_{jk}/n when there are n jobs in its queue, while server 0 is an *infinite* server, which serves all waiting jobs simultaneously, serving a job of type j at rate λ_j independent of how many jobs it is serving. A job of type of type j at a station $k \in \{0, \ldots, K\}$ waits to be at the front of the queue (if $k \geq 1$), then waits to be served, then moves to server l with probability $p_{j;kl}$. All service times are independent exponential random variables and the movement between servers is Markovian and independent of the service times.

The entire model is Markovian if one takes the state vector to be the $J \times (K+1)$ matrix $Q = Q(t)$ giving the number of jobs of each type waiting at each server. Let $n_j = n_j(t) = \sum_{k=0}^{K} Q_{jk}(t)$ be the number of jobs of type j present in the system at time t. The assumptions of our model (closed, with no class hopping) imply that the vector $\mathbf{n} = (n_1, \ldots, n_J)$ is constant. Assuming also that enough of the $p_{j;kl}$ are nonzero that any job at any server can get to any other server, the process started from any $\mathbf{n} = \mathbf{n}(0)$ will be an ergodic Markov chain on the set of states $S(\mathbf{n})$ of $J \times (K+1)$ matrices with row sums $\mathbf{n}$. There will be a unique stationary distribution $\pi_{\mathbf{n}}$. Explicit product formulae exist for $\pi_{\mathbf{n}}(Q)$ of the form $\pi_{\mathbf{n}}(Q) = \frac{1}{G(\mathbf{n})} P_{\mathbf{n}}(Q)$, where P is a large product and $G(\mathbf{n})$ is the normalizing constant or *partition function*. Typically, $P_{\mathbf{n}}$ is easier to estimate than $G(\mathbf{n})$, so much of the work in analyzing this sort of queuing network is in estimating $G(\mathbf{n})$.

It turns out that there is a relatively simple generating function for the quantities $G(\mathbf{n})$. One must first derive quantities ρ_{ji}, which are stationary probabilities for a single job migrating through the network. The *partition generating function* is then given by

$$F(\mathbf{z}) := \sum_{\mathbf{n}} G(\mathbf{n}) \mathbf{z}^{\mathbf{n}} = \frac{\exp(z_1 + \cdots + z_j)}{\prod_{i=1}^{K} \left(1 - \sum_{j=1}^{J} \rho_{ji} z_j\right)} . \tag{4.26}$$

This equation is given in [43, equation (44)] (though the definition of ρ_{ji} here is sketchy) or in [10, equation (2.26)] for a slightly more general model (more than one infinite server).

Evidently, the denominator of F decomposes into linear factors. A general analysis of this case, including an algorithm for identifying contrib, is provided in [4]. Here, we work the simplest nontrivial case, $J = K = 2$. Thus

$$F(x, y) = \frac{\exp(x + y)}{(1 - \rho_{11}\, x - \rho_{21}\, y)(1 - \rho_{12}\, x - \rho_{22}\, y)}$$

for positive constants ρ_{ji}. This example is worked in [10] but without the infinite server.

Assume that $\rho_{11} > \rho_{12}$, which loses no generality except for allowing $\rho_{11} = \rho_{12}$. The most interesting case is then $\rho_{22} > \rho_{21}$. The singular variety $\mathcal{V}$ consists of two lines with negative slopes, and we have assumed order relations implying that the lines intersect in the positive quadrant. This is exactly what is shown in Figure 3.

To recall the information from that figure, let

$$D := \rho_{11}\rho_{22} - \rho_{12}\rho_{21}$$

denote the determinant of the matrix $[\rho_{ji}]$. The set $\mathcal{V}$ has three strata, one being the point

$$(x_0, y_0) := \left(\frac{\rho_{22} - \rho_{21}}{D}, \frac{\rho_{11} - \rho_{12}}{D} \right),$$

where the lines meet, and the others being the two lines, each with the point (x_0, y_0) removed. Whenever a stratum is linear, there is a unique solution to the critical point equations (3.5) on the stratum (use concavity of the logarithm to see this), hence $\mathsf{crit}_{\overline{\mathbf{r}}}$ has cardinality 3 unless two of the critical points coincide. One of the critical points is always the point (x_0, y_0). Figure 10 shows the location of all three critical points as $\overline{\mathbf{r}}$ varies from near horizontal to near diagonal to near vertical.

The three configurations correspond to a division of the positive orthant of $\mathbb{RP}^1$ into three regions. Transitions between the regions occur when $\overline{\mathbf{r}}$ is equal to

$$\lambda_1 := \frac{\rho_{11}(\rho_{22} - \rho_{21})}{\rho_{21}(\rho_{11} - \rho_{12})}$$

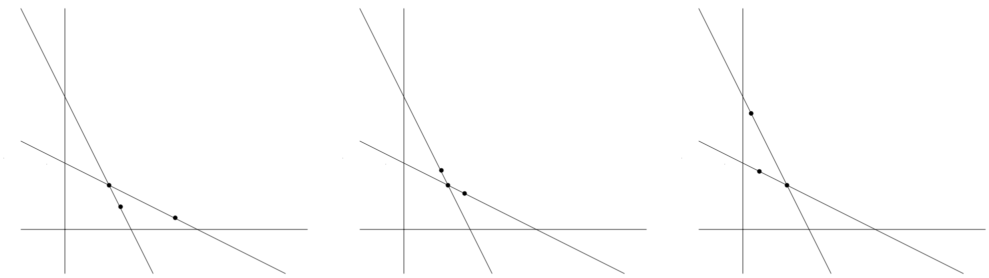

Fig. 10 *The three possible configurations of critical points.*

or

$$\lambda_2 := \frac{\rho_{12}(\rho_{22} - \rho_{21})}{\rho_{22}(\rho_{11} - \rho_{12})} .$$

The middle region is the $\mathbf{K}(x_0, y_0)$ of directions between λ_1 and λ_2. In the interior of this cone, asymptotics are given by Corollary 3.24. The determinant of the Hessian of the denominator of F, evaluated at (x_0, y_0), is equal to $-D^2$. The corollary therefore yields

$$a_{rs} \sim x_0^{-r} y_0^{-s} \frac{\exp(x_0 + y_0)}{x_0\, y_0\, D} .$$

To be completely explicit, consider the denominator from Figure 3. Here $\rho_{11} = \rho_{22} = 2/3$ and $\rho_{12} = \rho_{21} = 1/3$. The boundaries of the cone have slopes 1/2 and 2. Both x_0 and y_0 are equal to 1, so for $(r, s) \to \infty$ in a compact subcone of $\mathbf{K}(x_0, y_0)$,

$$a_{rs} = \frac{e^2}{1/3} = 3e^2 \approx 22.1671 .$$

The actual value for $r = s = 80$ is $a_{80,80} \approx 22.1668$ and the the relative error is 1.7×10^{-5}. Even at $r = s = 40$, the error is only one-quarter of one percent. As in section 4.4, there is a Gaussian scaling window at the boundary of $\mathbf{K}(x_0, y_0)$. Thus, for example, $a_{n,2n} \sim \frac{3}{2}e^2 \approx 11.083$. The error decays more slowly on the boundary (order $n^{-1/2}$, though our treatment here does not extend to this depth), so for instance $a_{40,80} \approx 10.893$ differs from the limiting value of $\frac{3}{2}e^2$ by 1.7%. The standard deviation in the scaling window is a little under $\sqrt{n}$; for example, when $n = 80$ the standard deviation is around 8. Checking $a_{48,80}$ we see a value of roughly 18.45, which is 83% of the interior value of $3e^2$, while at two standard deviations above we get $a_{56,80} \approx 21.49$ or 97% of the interior value; at one and two standard deviations below we have $a_{32,80} \approx 3.018$ and $a_{24,80} \approx 0.241$, which are, respectively, 13.6% and 1% of the interior value.

5. Transfer Matrices.

Suppose we have a class $\mathcal{C}$ of combinatorial objects that may be put in bijective correspondence with the collection of paths in a finite directed graph. For example, it is shown in [66, Example 4.7.7] that the class $\mathcal{C}$ of permutations $\pi \in \bigcup_{n=0}^{\infty} \mathcal{S}_n$ such that $|\pi(j) - j| \le 1$ for all j in the domain of π is of this type. Specifically, the digraph has vertices $V := \{-1, 0, 1\}^2 \setminus \{(0, -1), (1, 0)\}$, with a visit to the vertex (a, b) corresponding to j for which $\pi(j) = a$ and $\pi(j + 1) = b$; the vertices

$(0,-1)$ and $(1,0)$ do not occur because this would require $\pi(j) = \pi(j+1)$, which cannot happen; the edge set E is all elements of V^2 except those connecting $(1,x)$ to $(y,-1)$ for some x and y, since this would require $\pi(j) = \pi(j+2)$; the correspondence is complete because once $\pi(j), \pi(j+1)$, and $\pi(j+2)$ are distinct, no more collisions are possible.

The *transfer-matrix method* is a general device for producing a generating function which counts paths from a vertex i to a vertex j according to a vector weight function W such that $W(v_1,\ldots,v_n) = \sum_{j=1}^n w(v_i,v_j)$ for some function $w : V \times V \to \mathbb{N}^d$.

PROPOSITION 5.1 (Theorem 4.7.2 of [66]). *Let A be the weight matrix, that is, the matrix defined by $A_{ij} = \delta_{i,j}\mathbf{z}^{w(i,j)}$, where $\mathbf{z} = (z_1,\ldots,z_d)$ and $\delta_{i,j}$ denotes 1 if $(i,j) \in E$, and 0 otherwise. Let $\mathcal{C}_{ij}$ be the subclass of paths starting at i and ending at j and define $F(\mathbf{z}) = \sum_{\gamma\in\mathcal{C}_{ij}} \mathbf{z}^{W(\gamma)}$. Then*

$$F(\mathbf{z}) = \left((I-A)^{-1}\right)_{ij} = \frac{(-1)^{i+j}\det(I-A:i,j)}{\det(I-A)},$$

where $(M : i,j)$ denotes the (i,j)-cofactor of M, that is, M with row i and column j removed. ∎

Thus any class to which the transfer-matrix method applies will have a multivariate rational generating function. It is obvious, for example, that the transfer-matrix method applies to the class of words in a finite alphabet with certain transitions forbidden. The examples discussed in [66, pp. 241–260] range from restricted permutations to forbidden transition problems to the derivation of polyomino identities including (4.17), and finally to a very general result about counting classes that factor when viewed as monoids. Probably because techniques for extracting asymptotics were not widely known in the multivariate case, the discussion in [66] centers around counting by a single variable: in the end all weights are set equal to a single variable, x, reducing multivariate formulae such as (4.18) to univariate ones such as (4.17). But the methods in the present paper allow us to handle multivariate rational functions almost as easily as univariate functions, so we are able to derive joint asymptotics for several statistics at once, which is useful to the degree that we care about joint statistics.

5.1. Restricted Switching. Some of the examples that we have already seen, such as enumeration of Smirnov words, may be cast as transfer-matrix computations. A simpler example of the method is the following path counting problem.

Let G be the graph on $K+L+2$ vertices which is the union of two complete graphs of sizes $K+1$ and $L+1$ with a loop at every vertex and one edge $\overline{xy}$ between them. Paths on this graph correspond perhaps to a message or task being passed around two workgroups, with communication between the workgroups not allowed except between the bosses. If we sample uniformly among paths of length n, how much time does the message spend, say, among the common members of group 1 (excluding the boss)?

We can model this efficiently by a four-vertex graph, with vertices C_1, B_1, B_2, and C_2, where B stands for boss (so $x = B_1$ and $y = B_2$) and C refers to the workgroup. We have collapsed K vertices to form C_1 and L vertices to form C_2, so a path whose number of visits to C_1 is r and to C_2 is s will count for $K^r L^s$ actual paths.

Let $F(u,v,z) = \sum_\omega u^{N(\omega,C_1)} v^{N(\omega,C_2)} z^{|\omega|}$, where $N(\omega, C_j)$ denotes the number of visits by ω to C_j and $|\omega|$ is the length of the path ω. By Proposition 5.1, the function

F is rational with denominator $H = \det(I - A)$, where

$$A = \begin{bmatrix} uz & uz & 0 & 0 \\ z & z & z & 0 \\ 0 & z & z & z \\ 0 & 0 & vz & vz \end{bmatrix}$$

and the states are ordered C_1, B_1, B_2, C_2. Subtracting from the identity and taking the determinant yields

$$H = uz^2 + uz^2v - uz - uz^4v + z^2v - 2\,z - zv + 1 + z^3v + uz^3\,.$$

Giving each visit to C_1 weight K corresponds to the substitution $u = Ku'$ and, similarly, $v = Lv'$. Applying Theorem 3.27 and reframing the results in terms of the original u and v shows that the proportion of time a long path spends among C_1 tends as $n \to \infty$ to

$$\left.\frac{u\partial H/\partial u}{z\partial H/\partial z}\right|_{(K,L,z_0)},$$

where z_0 is the minimum modulus root of $H(K, L, z)$. In the present example, this yields

$$\left.\frac{K\left(-z - zL + 1 + z^3L - z^2\right)}{-2\,Kz - 2\,KzL + K + 4\,Kz^3L - 2\,zL + 2 + L - 3\,z^2L - 3\,Kz^2}\right|_{z=z_0}$$

for the proportion of time spent in C_1, where z_0 is the minimal modulus root of $H(K, L, z)$. Trying $K = 1$, $L = 1$ as an example, we find that $H(1, 1, z) = 1 - 4z + 2z^2 + 3z^3 - z^4$, leading to $z_0 \approx 0.381966$ and a proportion of just over 1/8 for the time spent in C_1. There are 4 vertices, so the portion of time the task spends at the isolated employee is just over half what it would have been had bosses and employees had equal access to communication. This effect is more marked when the workgroups have different sizes. Increasing the size of the second group to 2, we plug in $K = 1, L = 2$ and find that $z_0 \approx 0.311108$ and that the fraction of time spent in C_1 has plummeted to just under 0.024.

5.2. Connector Matrices. For a more substantial example, we have chosen a further refinement of the transfer-matrix method, called the *connector-matrix method.* The specific problem we will look at is the enumeration of sequences with forbidden substrings, enumerated by the composition of the sequence. This is a problem in which we are guaranteed (by the general transfer-matrix methodology) to get a multivariate rational generating function, but for which a more clever analysis greatly reduces the complexity of the computations. Our discussion of this method is distilled from Goulden and Jackson's exposition [32].

Let S be a finite alphabet and let T be a finite set of words (elements of $\bigcup_n S^n$). Let $\mathcal{C}$ be the class of words with no substrings in T, that is, those $(y_0, \ldots, y_n) \in \mathcal{C}$ such that $(y_j, \ldots, y_{j+k}) \notin T$ for all $j, k \geq 0$. We may reduce this to a forbidden transition problem with vertex set $V := S^k$ with k the maximum length of a word in T. This proves the generating function will be rational, but is a very unwieldy way to compute it, involving an $|S|^k$ by $|S|^k$ determinant.

The connector-matrix method of [35], as presented in [32, pp. 135–136], finds a much more efficient way to solve the forbidden substring problem (see also an elementary proof of these results via martingales [46]). These papers show that it is easy to enumerate sequences containing at least v_i occurrences of the forbidden substring ω_i. By inclusion-exclusion, one can then determine the number containing precisely y_i occurrences of ω_i, and then set $\mathbf{y} = \mathbf{0}$ to count sequences entirely avoiding the forbidden substrings.

To state the result, let $T = \{\omega_1, \ldots, \omega_m\}$ be a finite set of forbidden substrings in the alphabet $[d] := \{1, \ldots, d\}$. We wish to count words according to how many occurrences of each symbol $1, \ldots, d$ they contain and how many occurrences of each forbidden word they contain. Thus, for a word η, we define the weight $\tau(\eta)$ to be the d-vector counting the number of occurrences of each letter, and we let $\sigma(\eta)$ be the m-vector counting how many occurrences of each forbidden substring (possibly overlapping) occur in η. Let F be the $(d+m)$-variate generating function with variables $x_1, \ldots, x_d, y_1, \ldots, y_m$ defined by

$$F(\mathbf{x}, \mathbf{y}) := \sum_{\eta} \mathbf{x}^{\tau(\eta)} \mathbf{y}^{\sigma(\eta)} .$$

Proposition 5.2 (Theorem 2.8.6 and Lemma 2.8.10 of [32]). *Given a pair of finite words, ω and ω', let $\mathsf{connect}(\omega, \omega')$ denote the sum of the weight of α over all words α such that some initial segment β of ω' is equal to a final segment of ω and α is the initial unused segment of ω. Formally,*

$$\mathsf{connect}(\omega, \omega') = \sum_{\alpha : (\exists \beta, \gamma)\, \omega = \alpha\beta, \omega' = \beta\gamma} \tau(\alpha) .$$

Let $\mathbf{V}$ be the square matrix of size m defined by $\mathbf{V}_{ij} = \mathsf{connect}(\omega_i, \omega_j)$, define the diagonal matrices $\mathbf{Y} := \mathrm{diag}\,(y_1, \ldots, y_m)$, $\mathbf{L} := \mathrm{diag}\,(\mathbf{x}^{\tau(\omega_1)}, \ldots, \mathbf{x}^{\tau(\omega_m)})$, and let $\mathbf{J}$ be the m-by-m matrix of ones. Then

$$F(\mathbf{x}, \mathbf{y}) = [1 - (x_1 + \cdots + x_d) - C(\mathbf{x}, \mathbf{y} - \mathbf{1})]^{-1} ,$$

where

$$C(\mathbf{x}, \mathbf{y}) = \mathrm{trace}\, \left((\mathbf{I} - \mathbf{Y}\mathbf{V})^{-1} \mathbf{Y}\mathbf{L}\mathbf{J} \right) .$$

In particular, setting $\mathbf{y} = \mathbf{0}$, the generating function for the words with no occurrences of any forbidden substring is

$$F(\mathbf{x}, \mathbf{0}) = \left[1 - (x_1 + \cdots + x_d) + \mathrm{trace}\, \left((\mathbf{I} + \mathbf{V})^{-1} \mathbf{L}\mathbf{J} \right) \right]^{-1} . \tag{5.1}$$ □

Remark. The function C is the so-called cluster generating function, whose $(\mathbf{r}, \mathbf{s})$-coefficient counts strings of composition $\mathbf{r}$ once for each way that the collection of s_j occurrence of the substring ω_j may be found in the string.

5.3. Forbidden Substring Example. Goulden and Jackson [32] apply the results of the previous subsection to an example. Let $S = \{0, 1\}$ be the binary alphabet, and let $T = \{\omega_1, \omega_2\} = \{10101101, 1110101\}$. The final substrings of lengths 1 and 3 of ω_1 are initial substrings of ω_1 with corresponding leftover pieces 1010110 and 10101. Thus $V_{11} = x_1^3 x_2^4 + x_1^2 x_2^3$. Computing the other three entries similarly, we get

$$\mathbf{V} = \begin{bmatrix} x_1^3 x_2^4 + x_1^2 x_2^3 & x_1^3 x_2^4 \\ x_1^2 x_2^4 + x_1 x_2^3 + x_2^2 & x_1^2 x_2^4 \end{bmatrix} .$$

We also obtain

$$\mathbf{L} = \begin{bmatrix} x_1^3 x_2^5 & 0 \\ 0 & x_1^2 x_2^5 \end{bmatrix}.$$

Plugging the above results into (5.1), denoting (x_1, x_2) by (x, y), finally yields

$$F(x,y) = \frac{1 + x^2y^3 + x^2y^4 + x^3y^4 - x^3y^6}{1 - x - y + x^2y^3 - x^3y^3 - x^4y^4 - x^3y^6 + x^4y^6}. \tag{5.2}$$

We make the usual preliminary computations on F. Write G and H for the numerator and denominator in (5.2). A Gröbner basis computation in Maple quickly tells us that H has no singularities and that $H = G = 0$ does not occur in the positive quadrant of $\mathbb{R}^2$ (the last command finds real roots of the relevant polynomial):

```
> Basis([H , diff(H,x) , diff(H,y)] , plex(x,y));

                                        [1]
> gb:=Basis([G,H],plex(x,y));

                   4    3       2        3             2
                 [y  + y  - 2 y  + 1, y  + x - y + y  + 1]

> fsolve(gb[1], y);
                                  -1.905166168, -0.6710436067
```

Similarly, we see that Q and H vanish simultaneously only at $(1, 0)$ and at $(0, 1)$. Using our a priori knowledge that the coefficients of F are nonnegative, we apply Corollary 3.18 and look for minimal points on the lowest arc of the graph of $\mathcal{V}$ in the first quadrant. The plot of this is visually indistinguishable from the line segment $x + y = 1$, which is not surprising because the forbidden substrings only affect the terms of the generating function of order 7 and higher. More computer algebra shows the point $(x, y) \in \mathsf{contrib}_{\overline{\mathbf{r}}}$ to be algebraic of degree 21.

One question we might ask next is to what degree the forbidden substrings bias the typical word to contain an unequal number of zeros and ones. One might imagine there will be a slight preference for zeros since the forbidden substrings contain mostly ones. To find out the composition of the typical word, we apply Theorem 3.28. Setting $y = x$ gives the univariate generating function

$$f(x) = F(x,x) = \frac{1 + x^5 + x^6 + x^7 - x^9}{1 - 2x + x^5 - x^6 - x^8 - x^9 + x^{10}} = \sum_n N(n)x^n$$

for the number $N(n)$ of words of length n with no occurrences of forbidden substrings. The root of the denominator of minimum modulus is $x_0 = 0.505496\ldots$, whence $N(n) \sim C(1/x_0)^n$.

The direction associated to the point (x_0, x_0) is given by the projective point

$$\overline{\mathbf{r}} = \left(x\frac{\partial H}{\partial x}, y\frac{\partial H}{\partial y} \right)(x_0, x_0). \tag{5.3}$$

The ratio $\lambda_0 = r/s$ of zeros to ones for this $\overline{\mathbf{r}}$ is a rational function of x_0. We may evaluate x_0 numerically and plug it into this rational function, but the numerics will be more accurate if we reduce algebraically first in $\mathbb{Q}[x_0]$. Specifically, we may first reduce the rational function to a polynomial by inverting the denominator modulo the minimal polynomial for x_0 (Maple's `gcdex` function) then multiplying by the numerator and reducing again. Then, from the representation $\lambda_0 = P(x_0)$, we may produce the minimal polynomial for λ_0 by writing the powers of λ_0 all as polynomials of degree 9 in x_0 and determining the linear relation holding among the powers $\lambda_0^0, \ldots, \lambda_0^{10}$. We find in the end that $\lambda_0 = 1.059834\ldots$. Thus, indeed, there is a slight bias toward zeros.

Suppose we wish to know how long a string must be before the count of zeros and ones tells us whether the string was sampled from the measure avoiding 10101101 and 1110101 versus the uniform measure. We may answer this by means of the local central limit behavior described in Theorem 3.28. We may verify that the proportion of zeros is distributed as

$$\frac{\lambda_0}{\lambda_0+1} + cn^{-1/2} N(0,1),$$

where $N(0,1)$ denotes a standard normal. Once we have computed the constant c, we will know how big n must be before $\frac{\lambda_0}{\lambda_0+1} - \frac{1}{2} \gg cn^{-1/2}$ and the count of zeros and ones will tip us off as to which of the two measures we are seeing.

6. Lagrange Inversion. Suppose that a univariate generating function $f(z)$ satisfies the functional equation $f(z) = z\phi(f(z))$ for some function ϕ analytic at the origin and not vanishing there. Such functions often arise, among other places, in graph and tree enumeration problems. If ϕ is a polynomial, then f is algebraic, but even in this case it may not be possible to solve for f explicitly. A better way to analyze f is via the Lagrange inversion formula. One common formulation states [32, Theorem 1.2.4] that

$$[z^n]f(z) = \frac{1}{n}\left[y^{n-1}\right]\phi(y)^n, \tag{6.1}$$

where $[y^n]$ denotes the coefficient of y^n.

To evaluate the right-hand side of (6.1), we look at the generating function

$$\frac{1}{1 - x\phi(y)} = \sum_{n=0}^{\infty} x^n \phi(y)^n,$$

which generates the powers of ϕ. The $x^n y^{n-1}$ term of this sum is the same as the y^{n-1} term of $\phi(y)^n$. In other words,

$$[z^n]f(z) = \frac{1}{n}[x^n y^n]\frac{y}{1 - x\phi(y)}\,. \tag{6.2}$$

This is a special case of the more general formula for $\psi(f(z))$,

$$[z^n]\psi(f(z)) = \frac{1}{n}[x^n y^n]\frac{y\psi'(y)}{1 - x\phi(y)}\,. \tag{6.3}$$

These formulae hold at the level of formal power series, and, if ψ and ϕ have nonzero radius of convergence, at the level of analytic functions.

We can now apply the analysis leading to (4.10) (noting that the series $v(x)$ in that equation is presently called $\phi(y)$, while the series $\phi(x)$ there is here called $y\psi'(y)$). Assume ϕ has degree at least 2—the easy case $\phi(z) = az + b$ may be handled directly.

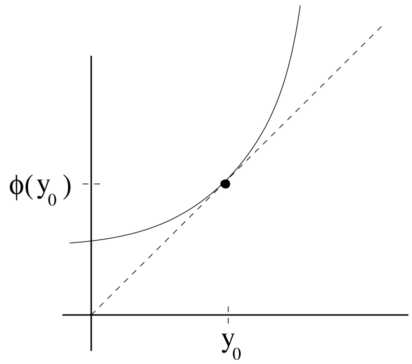

Fig. 11

Recall the definitions of μ and σ^2 from section 4.3. We are interested only in coefficients of $x^n y^n$, that is, the diagonal direction. Set $\mu(\phi; y) := y\phi'/\phi$ equal to 1; geometrically, we graph $\phi(y)$ against y and require that the secant line from the origin to the point $(y, \phi(y))$ be tangent to the graph there (see Figure 11). Letting y_0 denote a solution to this, we then have a point $(1/\phi(y_0), y_0)$ in the set crit_1 and at this point the quantity $x^{-n}y^{-n}$ is equal to $(\phi(y_0)/y_0)^n = \phi'(y_0)^n$. In (4.5) and (4.6) we have $\mu(\phi; y_0) = 1$ and consequently $\sigma^2(\phi; y_0)$ simplifies to $y_0^2\phi''(y_0)/\phi(y_0)$. Putting this together with the asymptotic formula (4.10), setting $r = s = n$, and simplifying leads to the following proposition. Note that it is easily shown that f is aperiodic if and only if ϕ is aperiodic.

PROPOSITION 6.1. *Let ϕ be analytic and nonvanishing at the origin, aperiodic with nonnegative coefficients, with degree at least 2 at infinity. Let f be the nonnegative series satisfying $f(z) = z\phi(f(z))$. Let $y = y_0$ be the positive solution of $y\phi'(y) = \phi(y)$. Then if ψ has radius of convergence strictly greater than y_0, we have*

$$[z^n]\psi(f(z)) \sim \phi'(y_0)^n n^{-3/2} \sum_{k\geq 0} b_k n^{-k}, \tag{6.4}$$

where $b_0 = \frac{y_0\psi'(y_0)}{\sqrt{2\pi\phi''(y_0)/\phi(y_0)}}$. ∎

A variant of Proposition 6.1, proved by other means, is found in [21, Theorem VI.6]. That result, proved by univariate methods, is stronger than Proposition 6.1 in some ways. For example, it can handle the estimation of $[z^n]T(z)(1 - T(z))^{-2}$, where $T(z) = z\exp(T(z))$, which occurs in the study of random mappings, whereas an attempt to use Proposition 6.1 directly runs into the problem that $y_0 = 1$ and $\psi(z) = (1 - z)^{-2}$ has radius of convergence 1.

6.1. Bivariate Asymptotics. We now discuss bivariate asymptotics. In the special case $\psi(y) = y^k$ for fixed k, the above results on Lagrange inversion yield the first order asymptotic

$$[z^n]f(z)^k \sim \frac{k}{n}\phi'(y_0)^n \frac{y_0^k}{\sqrt{2\pi n\phi''(y_0)/\phi(y_0)}},$$

where $y_0\phi'(y_0) = \phi'(y_0)$. We can also derive an asymptotic as both n and k approach ∞. We sketch an argument here (see [72] for details). The formula (4.10) supplies asymptotics whenever n/k belongs to a compact set of the interior of (A, B), where A is the order of f at 0 and B the order at ∞. In the present case we have $A = 1$ and $B = \infty$. We may then use the defining relation for $f(z)$ to express the formula in terms of ϕ only, leading to the following result.

PROPOSITION 6.2. *Let ϕ be analytic and nonvanishing at the origin, with nonnegative coefficients, aperiodic, and of order at least 2 at infinity. Let f be the positive series satisfying $f(z) = z\phi(f(z))$. For each n, k, set $\lambda = k/n$ and let $y = y_\lambda$ be the positive real solution of the equation $\mu(\phi; y) = (1 - \lambda)$. Then*

$$(6.5) \quad [z^n]f(z)^k \sim \lambda\phi(y_\lambda)^n y_\lambda^{k-n} \frac{1}{\sqrt{2\pi n\sigma^2(\phi; y_\lambda)}} = \lambda(1-\lambda)^{-n}\phi'(y_\lambda)^n \frac{y_\lambda^k}{\sqrt{2\pi n\sigma^2(\phi; y_\lambda)}}.$$

Here μ and σ^2 are given by (4.5) *and* (4.6)*, respectively. The asymptotic approximation holds uniformly, provided that λ lies in a compact subset of* $(0, 1)$. □

The similarity between (6.4) and (6.5) seems to indicate that a version of Proposition 6.2 that holds uniformly for $0 \le k/n \le 1 - \varepsilon$ should apply. This is consistent with a result of Drmota [18] but is seemingly more general and warrants further study. Such a result can very likely be obtained using the extension by Lladser [50] of the methods of [57], as can results of Meir and Moon [52] and Gardy [27] related to Proposition 6.2.

6.2. Trees in a Simple Variety. We now discuss a well-known situation [21, section VII.2] in which the foregoing results can be applied.

Consider the class of ordered (plane) unlabeled trees belonging to a so-called *simple variety*, namely, a class $\mathcal{W}$ defined by the restriction that each node may have a number of children belonging to a fixed subset Ω of $\mathbb{N}$. Some commonly used simple varieties are regular d-ary trees, $\Omega = \{0, d\}$; unary-binary trees, $\Omega = \{0, 1, 2\}$; and general plane trees, $\Omega = \mathbb{N}$. The generating function $f(z)$ counting trees by nodes satisfies $f(z) = z\omega(f(z))$, where $\omega(y) = \sum_{t\in\Omega} y^t$.

Provided that ω is aperiodic, Proposition 6.1 applies. The form of ω shows that the equation $y\omega'(y) = \omega(y)$ always has a unique solution strictly between 0 and 1. As a simple example, we compute the asymptotics for the number of general plane trees with n nodes (the exact answer being the Catalan number C_{n-1}). The equation $y\phi'(y) = \phi(y)$ has solution $y = 1/2$, corresponding to $\phi'(y) = 1/4$. Thus we obtain

$$[z^n]f(z) \sim 4^{n-1}\frac{1}{\sqrt{\pi n^3}}$$

in accordance with Stirling's approximation applied to the expression of C_{n-1} in terms of factorials.

Proposition 6.1 does not directly apply to the case of regular d-ary trees. Rather, we use a version where the contribution from several minimal points on the same torus must be added. The details are as follows. The equation $\mu(\phi; y) = 1$ has solutions $\omega\rho$, where $\rho = (d-1)^{-1/d}$ and $\omega^d = 1$. Each of these is a contributing critical point and the corresponding leading term asymptotic contribution is $C\omega^{(n-1)}\alpha^n n^{-3/2}$, where $C = \rho^2(2\pi(d-1))^{-1/2}$ and $\alpha^d = d^d/((d-1)^{d-1})$. Thus the asymptotic leading term is 0 unless d divides $n - 1$, in which case it is $dC\alpha^n n^{-3/2}$.

We note in passing that one can avoid the last computation as follows [21, section I.5, Example 13]. There is a bijection between the class of d-ary trees and the class $\mathcal{C}$ of trees with vertices of degree at most d but allowing for $\binom{d}{j}$ types of nodes of degree j. The pruning map removes all external nodes (nodes of degree 0) from a d-ary tree, then labels each node of the pruned tree with the set of children that were removed. Pruning always maps a tree of $1+dm$ nodes to a tree of m nodes. The extension of the above asymptotics to a multiset of degrees (allowing for different types of children) is straightforward. Thus we can compute using the degree enumerator $g(z) = (1+z)^d$ of $\mathcal{C}$ in Proposition 6.1 and $[z^n]f(z) = [z^m]g(z)$.

Now we consider the *mean degree profile* of trees in $\mathcal{W}$. Let $\xi_k(t)$ be the number of nodes of degree k in the tree t, and $|t|$ the total number of nodes in t. A standard calculation [21, section VII.2.2, Example 5] shows that the cumulative generating function is

$$\sum_{t \in \mathcal{W}} z^{|t|} \xi_k(t) = z^2 \phi_k f(z)^{k-1} \phi'(f(z)),$$

where $\phi_k = [y^k]\phi(y)$. Thus we have

$$F(z,u) := \sum_{k \geq 0} \sum_{t \in \mathcal{W}} \xi_k(t) z^{|t|} u^k = \mu(f; z) z \phi(u f(z)) \,.$$

The mean number of nodes of degree k in a uniformly randomly chosen tree of size n from $\mathcal{W}$ is then given by

$$M_{nk} = \frac{[z^n u^k] F(z,u)}{[z^n] f(z)}.$$

Consider again the simplest case, general plane trees, with $\phi(y) = (1-y)^{-1}$. Then $F(z,u)$ corresponds to a Riordan array. A simple variant of Proposition 6.2 shows that

$$[z^n u^k] F(z,u) \sim y^k \phi'(y)^{n-1} \frac{1}{\sqrt{2\pi n \sigma^2(\phi; y)}},$$

where $\mu(\phi; y) = 1 - k/n$. This is easily solved to obtain $y = (n-k)/(2n-k)$ and hence we obtain routinely

$$M_{nk} \sim n \left(\frac{2n-k}{2n} \right)^{2n-2} \left(\frac{n-k}{2n-k} \right)^k \sqrt{\frac{n^2}{2(n-k)(2n-k)}} \tag{6.6}$$

uniformly as long as k/n is in a compact subset of $(0,1)$. The mean number of leaves (nodes of degree 0) in such a tree is well known [21, section III.5, Example 12] to be $n/2$, which is obtained by substituting $k=0$ in the right-hand side of (6.6). Thus it again seems likely that the approximation is in fact uniform on $[0, 1-\varepsilon]$.

6.3. Cores of Planar Graphs. A *rooted planar map* is a graph with a distinguished edge (the root) that can be embedded in the plane. A rooted planar map is completely specified by a planar graph, a distinguished edge, and a cyclic ordering of edges around each vertex. The *core* of a map (henceforth always a rooted planar map) is the largest 2-connected subgraph containing the root edge. The problem of the typical core size (cardinality of the core) of a map was considered in [25, 3]. The author's obtained a functional equation for the generating function

$$M(u,z) = \sum_{n,k} a_{n,k} z^n u^k,$$

where $a_{n,k}$ is the number of maps with n edges and core size k (that is, the core has cardinality k). They are interested in computing the probability distribution of the core size of a map sampled uniformly from among all maps of with n edges. Applying the general inversion formula (6.3) to their functional equations, they are able to compute

$$p(n,k) := \frac{a_{n,k}}{\sum_j a_{n,j}} = \frac{k}{n}[z^{n-1}]\psi'(z)\psi(z)^{k-1}\phi(z)^n, \tag{6.7}$$

where $\psi(z) = (z/3)(1-z/3)^2$ and $\phi(z) = 3(1+z)^2$ are the operators that arise in the functional equation for $M(u,z)$.

The mean size of the core was computed in [25]. For this it sufficed to prove a specialized result about coefficients of powers of functions asymptotic to $(1-z)^{-3/2}$ as $z \to 1$. The problem of finding a limit law for the distribution about the mean was taken up in [3]. One may arrive directly at an asymptotic formula for $p(n,k)$ if one rewrites (6.7) as

$$p(n,k) = \frac{k}{n}[x^k y^n z^n]\frac{xz\psi'(z)}{(1-x\psi(z))(1-y\phi(z))} \,.$$

The analysis of the resulting trivariate generating function is, however, quite challenging. In particular, there is a point where Q vanishes; a generalization of Theorem 3.19 [57, Theorem 3.3] tells us the asymptotics in this precise direction, but does not answer the more interesting question of asymptotics in a scaling window near that direction, which turns out to be $k = n/3 \pm n^{-2/3}$. An answer, in the framework of [57], is given in the doctoral work of Lladser [50, section 5.5]. A complete answer is given in [3] by reductions to a one-variable generating function, on which a coalescing saddle approximation is used. We will not go into details here.

7. The Kernel Method.

The kernel method is a means of producing a generating function for an array $\{a_{\mathbf{r}} : \mathbf{r} \in \mathbb{N}^d\}$ of numbers satisfying a linear recurrence

$$a_{\mathbf{r}} = \sum_{\mathbf{s}\in E} c_{\mathbf{s}} a_{\mathbf{r}-\mathbf{s}} \,. \tag{7.1}$$

Here E is a finite subset of $\mathbb{Z}^d \setminus \{0\}$ which is not necessarily a subset of $\mathbb{N}^d$ but whose convex hull must not intersect the closed negative orthant. The numbers $\{c_{\mathbf{s}} : s \in E\}$ are constants and the relation (7.1) holds for all $\mathbf{r}$ except those in the *boundary condition*, which will be made precise below. As usual, let $F(\mathbf{z}) = \sum a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$. In one variable F is always a rational function, but in more than one variable the generating function can be rational, algebraic, D-finite, or differentially transcendental (not D-finite). A classification along these lines, determined more or less by the number of coordinates in which points of E can be negative, is given in [11], which is a very nice exposition of the kernel method at an elementary level.

We are interested in the kernel method because it often produces generating functions to which Theorem 3.19 may be applied. Because the method is not all that well known, we include a detailed description, drawing heavily on [11]. We begin, though, with an example.

7.1. A Random Walk Problem.

Two players move their tokens toward the finish square, flipping a fair coin each time to see who moves forward one square. At present the distances to the finish are $1+r$ and $1+r+s$. If the second player passes the first player, the second player wins; if the first player reaches the finish square, the first player wins; if both players are on the square before the finish square, it is a draw. What is the probability of a draw?

Let a_{rs} be the probability of a draw, starting with initial positions $1 + r$ and $1 + r + s$. Conditioning on which player moves first, one finds the recursion

$$a_{rs} = \frac{a_{r,s-1} + a_{r-1,s+1}}{2},$$

which is valid for all $(r, s) \geq (0, 0)$ except for $(0, 0)$, provided that we define a_{rs} to be zero when one or more coordinates is negative. The relation $a_{rs} - (1/2)a_{r,s-1} - (1/2)a_{r-1,s+1} = 0$ suggests we multiply the generating function $F(x, y) := \sum a_{rs}x^r y^s$ by $1-(1/2)y-(1/2)(x/y)$. To clear denominators, we multiply by $2y$: define $Q(x, y) = 2y - y^2 - x$ and compute $Q \cdot F$. We see that the coefficients of this vanish with two exceptions: the x^0y^1 coefficient corresponds to $2a_{0,0} - a_{0,-1} - a_{-1,1}$, which is equal to 2, not 0, because the recursion does not hold at $(0, 0)$ (a_{00} is set equal to 1); the y^0x^j coefficients do not vanish for $j \geq 1$ because, due to clearing the denominator, these correspond to $2a_{j,-1} - a_{j,-2} - a_{j-1,0}$. This expression is nonzero since, by definition, only the third term is nonzero, but the value of the expression is not given by prescribed boundary conditions. That is, we have

$$Q(x, y)F(x, y) = 2y - h(x), \tag{7.2}$$

where $h(x) = \sum_{j\geq 1} a_{j-1,0}x^j = xF(x, 0)$ will not be known until we solve for F.

This generating function is in fact a simpler variant of the one derived in [44] for the waiting time until the two players collide, which is needed in the analysis of a sorting algorithm. Their solution is to observe that there is an analytic curve in a neighborhood of the origin on which Q vanishes. Solving $Q = 0$ for y in fact yields two solutions, one of which, $y = \xi(x) := 1 - \sqrt{1 - x}$, vanishes at the origin. Since ξ has a positive radius of convergence, we have, at the level of formal power series, that $Q(x, \xi(x)) = 0$, and substituting $\xi(x)$ for y in (7.2) gives

$$0 = Q(x, \xi(x))F(x, \xi(x)) = 2\xi(x) - h(x)\,.$$

Thus $h(x) = 2\xi(x)$ and

$$F(x, y) = 2\frac{y - \xi(x)}{Q(x, y)} = \frac{1}{1 + \sqrt{1 - x} - y}\,.$$

As is typical of the kernel method, the generating function F has a pole along the branch of the kernel variety $\{Q(x, y) = 0\}$ that does not pass through the origin. The function F is not meromorphic everywhere, having a branch singularity on the line $x = 1$, but it is meromorphic in neighborhoods of polydisks $\mathbf{D}(x, y)$ for minimal points $(x, 1 + \sqrt{1 - x})$ on the pole variety for $0 < x < 1$. For $0 < x < 1$, $\mathsf{dir}(x, 1 + \sqrt{1 - x}) = (2\sqrt{1 - x} + 2 - 2x)/x$. If we set this equal to λ and solve for x we find $x = 4(1 + \lambda)/(2 + \lambda)^2$ and $1 + \sqrt{1 - x} = (2 + 2\lambda)/(2 + \lambda)$. In other words,

$$\mathsf{contrib}_{\overline{\mathbf{r}}} = \left(\frac{4(1 + \lambda)}{(2 + \lambda)^2}, \frac{2(1 + \lambda)}{2 + \lambda}\right),$$

where $\lambda = s/r$. Plugging this into (3.6) gives

$$\begin{aligned} a_{rs} &\sim C(r + s)^{-1/2} \left(\frac{4r(r + s)}{(2r + s)^2}\right)^{-r} \left(\frac{2(r + s)}{2r + s}\right)^{-s} \\ &= \frac{C}{2^{2r+s}} \frac{(2r + s)^{2r+s}}{r^r(r + s)^{(r+s)}}\,. \end{aligned}$$

One recognizes in this formula the asymptotics of the binomial coefficient $\binom{2r+s}{r}$, and indeed the binomial coefficient may be obtained via a combinatorial analysis of the random walk paths.

7.2. Explanation of the Kernel Method. Because of the applicability of Theorem 3.19 to generating functions derived from the kernel method, we now give a short explanation of this method. We adopt the notation from the first paragraph of this section.

Let $\mathbf{p}$ be the coordinatewise infimum of points in $E \cup \{\mathbf{0}\}$, that is, the greatest element of $\mathbb{Z}^d$ such that $\mathbf{p} \le \mathbf{s}$ for every $\mathbf{s} \in E \cup \{\mathbf{0}\}$. Let

$$Q(\mathbf{z}) := \mathbf{z}^{-\mathbf{p}} \left(1 - \sum_{\mathbf{s} \in E} c_{\mathbf{s}} \mathbf{z}^{\mathbf{s}} \right),$$

where the normalization by $\mathbf{z}^{-\mathbf{p}}$ guarantees that Q is a polynomial but not divisible by any z_j. Partition $\mathbb{N}^d$ into two sets, Z and B, and assume that the relation (7.1) holds for all $\mathbf{r} \in Z$; the set B must be closed under $\le$ and the values $\{a_{\mathbf{r}} : \mathbf{r} \in B\}$ are specified explicitly by constants $\{b_{\mathbf{r}} : \mathbf{r} \in B\}$ rather than by (7.1).

If $E \subseteq \mathbb{N}^d$, then $\mathbf{p} = \mathbf{0}$, B can be arbitrary, F is a function of the form G/Q, and G is rational if the boundary conditions are rational. The analysis in this case is straightforward and the kernel method yields only what may be derived directly from the recursion for $a_{\mathbf{r}}$ in terms of $\{a_{\mathbf{s}} : \mathbf{s} \le \mathbf{r}\}$. For examples of this, see sections 4.1 and 4.2. We concentrate instead on the case where $d = 2$ and the second coordinate of points in E may be negative. It is known in this case [11, Theorem 13] that if the generating function for the boundary conditions is algebraic, then F is algebraic. On the other hand, we shall see that an outcome of the kernel method is that F will have a pole variety and will usually satisfy the meromorphicity condition in the remark after Theorem 3.16.

To apply the kernel method, one examines the product QF_Z, where for convenience we have let $F_Z := \sum_{\mathbf{r} \in Z} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ be the generating function for those values for which the recursion (7.1) holds. If we assume that the generating function $F_B := \sum_{\mathbf{r} \in B} b_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ for the prescribed boundary conditions is known, then since $F = F_Z + F_B$, finding F_Z is equivalent to finding F.

There are two kinds of contribution to QF_Z. First, for every pair $(\mathbf{r}, \mathbf{s})$ with $\mathbf{s} \in E$, $\mathbf{r} \in Z$, and $\mathbf{r} - \mathbf{s} \in B$, there is a term $c_{\mathbf{s}} b_{\mathbf{r}-\mathbf{s}} \mathbf{z}^{\mathbf{r}-\mathbf{p}}$; the corresponding coefficient of QF vanishes, but F_Z has been stripped of the boundary terms that cause the cancellation. Let

$$K(\mathbf{z}) := \sum_{\mathbf{r} \in Z, \mathbf{s} \in E, \mathbf{r}-\mathbf{s} \in B} c_{\mathbf{s}} b_{\mathbf{r}-\mathbf{s}} \mathbf{z}^{\mathbf{r}-\mathbf{p}}$$

denote the sum of these terms. The "K" stands for "known," because the coefficients of K are determined by the boundary conditions, which are known. Second, for every pair $(\mathbf{r}, \mathbf{s})$ with $\mathbf{s} \in E, \mathbf{r} - \mathbf{s} \in B$, and $\mathbf{r} \notin Z$, there is a term $-c_{\mathbf{s}} a_{\mathbf{r}-\mathbf{s}} \mathbf{z}^{\mathbf{r}-\mathbf{p}}$ due to the fact that the recursion does not hold at $\mathbf{r}$; let

$$U(\mathbf{z}) = \sum_{\mathbf{r}-\mathbf{s} \in Z, \mathbf{s} \in E, \mathbf{r} \notin Z} c_{\mathbf{s}} a_{\mathbf{r}-\mathbf{s}} \mathbf{z}^{\mathbf{r}-\mathbf{p}}$$

denote these terms. The "U" stands for "unknown," because these coefficients are not explicitly determined from the boundary conditions. It is not hard to show that

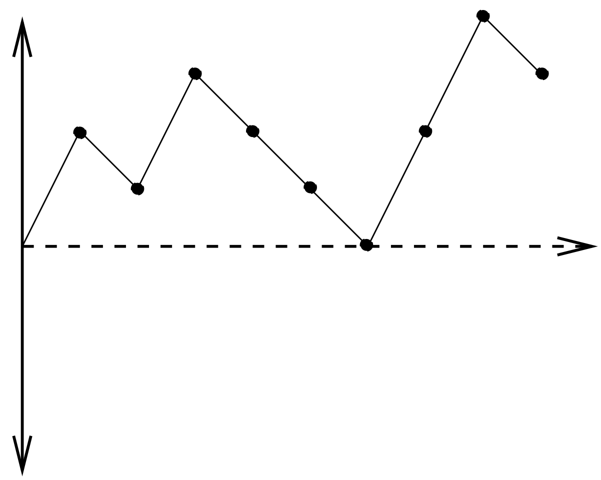

Fig. 12 *A generalized Dyck path of length nine with* $E = \{(1,2),(1,-1)\}$.

for any dimension d and any E whose convex hull does not contain a neighborhood of the origin, the following result holds.

PROPOSITION 7.1 (see [11, Theorem 5]). *Let* E, $\{c_{\mathbf{s}}\}$, $\mathbf{p}$, B, Z, $\{b_{\mathbf{s}} : \mathbf{s} \in B\}$, F_Z, *and* K *be as above. Then there is a unique set of values* $\{a_{\mathbf{r}} : \mathbf{r} \in Z\}$ *such that* (7.1) *holds for all* $\mathbf{r} \in Z$. *Consequently, there is a unique pair* (F, U) *of formal power series such that*

$$QF_Z = K - U\,. \qquad \square$$

In other words, the unknown power series U is determined from the data along with F_Z. Another way of thinking about this is that F_Z is trying to be the power series K/Q, but since Q vanishes at the origin, one must subtract some terms from K to cancel the function Q_0, where $Q = Q_0 Q_1$ and Q_0 consists of the branches of Q passing through the origin. The kernel method, as presented in [11], turns this intuition into a precise statement.

PROPOSITION 7.2 (see [11, equation (24)]). *Suppose* $d = 2$, $\mathbf{p} = (0, -p)$, *and* $F_B = 1$. *There will be exactly* p *formal power series* $\xi_1, \ldots, \xi_p$ *such that* $\xi_j(0) = 0$ *and* $Q(x, \xi_j(x)) = 0$, *and we may write* $Q(x,y) = -C(x)\prod_{j=1}^{p}(y - \xi_j(x))\prod_{j=1}^{r}(y - \rho_j(x))$ *for some* r *and* $\rho_1, \ldots, \rho_r$. *The generating function* F_Z *will then be given by*

$$F_Z(x,y) = \frac{K(x,y) - U(x,y)}{Q(x,y)} = \frac{\prod_{j=1}^{p}(y - \xi_j(x))}{Q(x,y)} = \frac{1}{-C(x)\prod_{j=1}^{r}(y - \rho_j(x))}\,. \qquad \square$$

We turn to some examples.

7.3. Dyck, Motzkin, Schröder, and Generalized Dyck Paths. Let E be a set $\{(r_1, s_1), \ldots, (r_k, s_k)\}$ of integer vectors with $r_j > 0$ for all j and $\min_j s_j = -p < 0 < \max_j s_j = P$. The generalized Dyck paths with increments in E to the point (r, s) in the first quadrant are the paths from $(0,0)$ to (r,s), with increments in E, which never go below the horizontal axis (see Figure 12; the sets E for three well-known special cases are shown in Figure 13).

Let $F(x,y) = \sum_{r,s} a_{rs} x^r y^s$, where a_{rs} is the number of generalized Dyck paths to the point (r,s). We will have $B = \{(0,0)\}$, $F_B = 1$, $Q(x,y) = y^p(1 - \sum_i x^{r_i} y^{s_i})$,

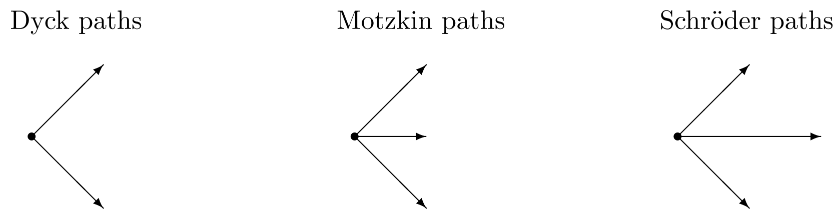

Fig. 13 *Legal steps for three types of paths.*

and $C(x) = \sum_{i:s_i=P} x^{r_i}$. We now discuss three well-known instances of such paths taken from [11].

Dyck Paths. When $E = \{(1,1),(1,-1)\}$, we have the original Dyck paths. We have $Q(x,y) = y - xy^2 - x$. Here $C(x) = x$ and $Q(x,y) = -x(y-\xi(x))(y-\rho(x))$, where $\xi(x) = (1-\sqrt{1-4x^2})/(2x)$ and $\rho(x) = (1+\sqrt{1-4x^2})/(2x)$ is the algebraic conjugate of ξ. Note that ρ is a formal Laurent series and $\rho\xi = 1$.

Thus we have, following the discussion above,

$$F(x,y) = \frac{1}{-x(y-\rho(x))} = \frac{\xi(x)/x}{1-y\xi(x)}.$$

Setting $y = 0$ recovers the fact that the Dyck paths coming back to the x-axis at $(2n,0)$ are counted by the Catalan number C_n.

Asymptotics are readily obtained using either the explicit or implicit form of ξ (noting the periodicity of ξ). Let us use the implicit form in this example, since we will be illustrating use of the explicit form below in the case of Schröder paths, where Lagrange inversion does not apply. The vanishing of $Q(x,y) = y - xy^2 - x$ occurs when $y = x\phi(y)$ with $\phi(y) = 1 + y^2$. Thus $\xi = x\phi(\xi)$. We use a slight variant of (6.5), namely,

$$[z^n]\xi(z)^{k+1}/z \sim \lambda\phi(y_\lambda)^{n+1} y_\lambda^{k-n} \frac{1}{\sqrt{2\pi n\sigma^2(\phi;y_\lambda)}},$$

where $\lambda = k/n$ and $\mu(\phi;y_\lambda) = 1-\lambda$.

The chain rule yields $\mu(\phi;y) = y\phi_y/\phi = y\phi_t t_y/\phi = (yt_y/t)(t\phi_t/\phi) = \mu(y^2;y)\mu(1+t;t)$ with $t = y^2$. Thus we solve $1-\lambda = 2t/(1+t)$ or $y^2 = (1-\lambda)/(1+\lambda) = (r-s)/(r+s)$. The two contributing points y_λ cancel out if $r-s$ is odd and reinforce if $r-s$ is even. We compute $\phi(y_\lambda) = 2/(1+\lambda) = 2r/(r+s)$ and $\sigma^2(\phi;y_\lambda) = 1-\lambda^2 = (r^2-s^2)/r^2$. We obtain

$$\begin{aligned} a_{rs} &\sim 2\,\frac{s}{r}\left(\frac{2r}{r+s}\right)^{r+1}\left(\frac{r-s}{r+s}\right)^{\frac{s-r}{2}} \frac{\sqrt{r}}{\sqrt{2\pi(r-s)(r+s)}} \\ &= \frac{2s}{(r+s)}\,\frac{r^r}{(\frac{r+s}{2})^{\frac{r+s}{2}}(\frac{r-s}{2})^{\frac{r-s}{2}}}\,\frac{\sqrt{r}}{\sqrt{2\pi(\frac{r-s}{2})(\frac{r+s}{2})}}, \end{aligned}$$

provided $r-s$ is even, and 0 otherwise. This is uniform for $0 < \delta \le s/r \le 1-\varepsilon < 1$.

Of course, in this simple example Lagrange inversion also gives an exact formula involving binomial coefficients, namely,

(7.3)
$$a_{rs} = [x^{r+1}]\xi(x)^{s+1} = \frac{s+1}{r+1}[y^{r-s}](1+y^2)^{r+1} = \frac{s+1}{r+1}\binom{r+1}{(r-s)/2} = \frac{2(s+1)}{r+s+2}\,\frac{r!}{(\frac{r-s}{2})!(\frac{r+s}{2})!}$$

when $r - s$ is even, and 0 otherwise. If we had instead computed $[z^{r+1}]\xi(z)^{s+1}$ using (6.5), we would have obtained a correct leading order asymptotic of a slightly different form (r, s replaced by $r+1, s+1$ in some places). In each case these asymptotics are consistent with what is obtained by applying Stirling's formula to the factorials in (7.3).

Motzkin Paths. Let $E = \{(1,1), (1,0), (1,-1)\}$. In this case the generalized Dyck paths are known as *Motzkin paths*. Again we have the case $Q(x, y) = y - xy^2 - x - xy$. Now ρ and ξ are given by $(1 - x \pm \sqrt{1 - 2x - 3x^2})/(2x)$ and again

$$F(x, y) = \frac{\xi(x)/x}{1 - y\xi(x)} = \frac{2}{1 - x + \sqrt{1 - 2x - 3x^2} - 2xy}\,.$$

This time ξ is given implicitly by $\xi = x(1+\xi+\xi^2)$ and the coefficients are not binomial coefficients, but the asymptotics are no harder to compute. Here, with $\lambda = s/r$, we have that $\mathsf{contrib}_{1-\lambda}$ is a singleton $\{y_\lambda\}$ by aperiodicity. The critical point equation is

$$(1 + \lambda)y^2 + \lambda y - (1 - \lambda) = 0,$$

and the solution is

$$y_\lambda = \frac{\sqrt{4 - 3\lambda^2} - \lambda}{2(1 + \lambda)}\,.$$

The minimal polynomial for $\sigma^2(\phi; y_\lambda)$ is found as in earlier sections to be

$$3S^2 + (6\lambda^2 + 12\lambda - 2)S + 3\lambda^4 - 24\lambda^3 + 65\lambda^2 - 68\lambda + 24.$$

This polynomial has two positive solutions for S for each given λ. The correct one is found by noting that σ^2 approaches 0 as λ approaches 0.

Schröder Paths. Here $E = \{(1,1), (2,0), (1,-1)\}$. We have $C(x) = x, Q(x, y) = y - xy^2 - x^2y - x$, and ρ and ξ are given by $(1 - x^2 \pm \sqrt{1 - 6x^2 + x^4})/(2x)$. This time Lagrange inversion does not obviously apply. We perform an explicit computation, noting the periodicity of F. We have $\mathsf{dir}(x, y) = \bar{\mathbf{r}}$ with

$$\lambda := \frac{s}{r} = \frac{\sqrt{1 - 6x^2 + x^4}}{1 + x^2}.$$

Then λ decreases from 1 to 0 as x increases from 0 to the smaller positive root of $1 - 6x^2 + x^4$, namely, $\sqrt{2} - 1$. We also have

$$x^2 = \frac{3 + \lambda^2 - 2\sqrt{(2 + 2\lambda^2)}}{1 - \lambda^2}.$$

Choosing the positive value of x, we see that asymptotics are given by

$$a_{rs} \sim 2Cx^{-r}y^{-s}s^{-1/2},$$

where $y = 1/\xi(x)$, when $r + s$ is even, and 0 otherwise. Any particular diagonal (with a value of λ between 0 and 1) can be extracted easily. For example, with $\lambda = 1/3$, we obtain $a_{3s,s} \sim 2C\gamma^s s^{-1/2}$, with $x = (3 - \sqrt{5})/2, y = (1 + \sqrt{5})/2$, $\gamma = (11 + 5\sqrt{5})/2 \approx 11.09016992$, and $C \approx 0.1526195310$. For $s = 12$ this approximation underestimates by about 2.9%, and for $s = 24$ by about 1.5%.

7.4. Pebble Configurations. Chung et al. [12] consider the following problem. Pebbles are placed on the nonnegative integer points of the plane. The pebble at (i, j) may be replaced by two pebbles, one at $(i+1, j)$ and one at $(i, j+1)$, provided this does not cause two pebbles to occupy the same point. Starting from a single pebble at the origin, it is known to be impossible to move all pebbles to infinity; in fact it is impossible to clear the region $\{(i, j) : 1 \leq i + j \leq 2\}$ [12, Lemma 2].

They consider the problem of enumerating minimal unavoidable configurations. More specifically, say that a set T is a *minimal unavoidable configuration* with respect to some starting configuration S if it is impossible starting from S to move all pebbles off T, but pebbles may be cleared from any proper subset of T. Let S_t denote the starting configuration where (i, j) is occupied if and only if $i+j = t$. Let $f_t(k)$ denote the number of sets in the region $\{(i, j) : i, j \geq 0; i + j \geq t + 1\}$ that are minimal unavoidable configurations for the starting configuration S_t.

Chung et al. [12] derive the recurrence

$$f_t(k) = f_{t-1}(k) + 2f_t(k-1) + f_{t+1}(k-2),$$

which holds whenever $t \geq 3$ and $k \geq 2$. Let

$$F(x, y) = \sum_{t,k\geq 0} f_t(k)x^k y^t .$$

According to the kernel method, we will have $F = \eta/Q$ for some η vanishing on the zero set of Q near the origin, where $Q(x, y) = x - (x + y)^2$. Using some more identities, Chung et al. are able to evaluate F explicitly. They state that they are primarily interested in $f_0(k)$, so they specialize to $F(x, 0)$ and compute the univariate asymptotics. It seems to us that the values $f_t(k)$ are of comparable interest, and we pursue asymptotics of the full generating function.

The formula for F is cumbersome, but its principal features are (i) a denominator of $P \cdot Q$, where P is the univariate polynomial $1 - 7x + 14x^2 - 9x^3$, and (ii) F is algebraic of degree 2 and is in $\mathcal{C}[x, y][\sqrt{1 - 4x}]$. The minimum modulus root of P is $x_0 \approx 0.2410859\ldots$. The algebraic singularity of F occurs along the line $x = 1/4$, so, conveniently, the branching is completely outside the closure of the domain of convergence and the meromorphicity assumption of the remark following Theorem 3.16 is satisfied.

The boundary of the domain of convergence in the first quadrant is composed of pieces of two curves, namely, $x = x_0$ and $y = (x + y)^2$. These intersect at the point (x_0, y_0), where $y_0 = (1 - 2x + \sqrt{1 - 4x})/2 \approx 0.3533286$ (see Figure 14). We are in the combinatorial case, so we know that minimal points will be found along these curves. Along the curve $y = (x + y)^2$ the direction $\mathsf{dir}(x, y)$ is given by $\lambda = r/s = (2 - 2\sqrt{y})/(2\sqrt{y} - 1)$. As (x, y) travels from $(0, 1)$ down the curve to (x_0, y_0), λ increases from 0 to $\lambda_0 \approx 4.295798\ldots$. At the point (x_0, y_0), the cone $\mathbf{K}(x_0, y_0)$ is the convex hull of the positive x-direction and the direction $(\lambda_0, 1)$.

We then have two sorts of asymptotics for $f_t(k)$. When $t < k/\lambda_0$, the asymptotics are given by Corollary 3.25. In this case we may evaluate $G(x_0, y_0) \approx 0.00154376$ and $\sqrt{-x_0^2 y_0^2 \mathcal{H}} \approx 0.02925688$ so that

$$f_t(k) \sim C x_0^{-k} y_0^{-t} \quad \text{with } C = 0.05276\ldots$$

uniformly as t/k varies over compact subsets of $(0, 1/\lambda_0)$. It is interesting to compare this to the asymptotics for $f_0(k)$. Setting $y = 0$ gives the univariate generating

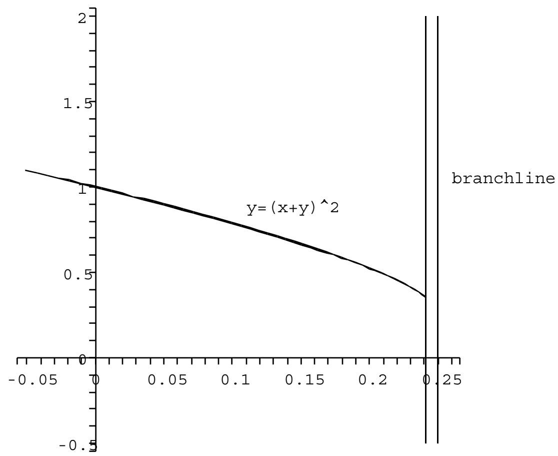

Fig. 14 *The domain of convergence and the algebraic singularity.*

function [12]

$$f(x) = \sum a_n x^n = x^2 \frac{(1-4x)^{1/2}(1-3x+x^2) - 1 + 5x - x^2 - 6x^3}{P(x)} .$$

We may compute

$$\lim_{n\to\infty} a_n x_0^k = \lim_{x\to x_0} (x_0 - x) f(x)$$

and find that

$$f_0(k) \sim c x_0^{-k} \quad \text{with } c = 0.016762\ldots.$$

In fact, one may calculate $\lim f_t(k) x_0^k$ for any fixed t by computing

$$f_t(x) := (t!)^{-1} \left(\frac{\partial}{\partial y}\right)^t F(x,0) \tag{7.4}$$

and again computing $c(t) := \lim_{x\to x_0}(x_0 - x) f_t(x)$. The pole at $x = x_0$ in F is removable: replacing the denominator PQ of F by $(P/(x-x_0))Q$ we see that $g(y) := \lim_{x\to x_0}(x_0 - x)F(x,y)$ has a simple pole at $y = y_0$ and that $\lim_{t\to\infty} y_0^t c(t) = \lim_{y\to y_0}(y_0 - y)g(y)$ is equal to C. In other words, we see that the asymptotics known to hold uniformly as t/k varies over compact subsets of $(0, 1/\lambda_0)$ actually hold over $[0, 1/\lambda_0)$ as long as $t \to \infty$, while for $k \to \infty$ with t fixed we use (7.4).

On the other hand, when $t/k > 1/\lambda_0$ we may solve for x and y to get

$$\mathsf{contrib}_{\overline{\mathsf{F}}} = (x,y) := \left\{ \left(\frac{k(2t+k)}{(2t+2r)^2} , \frac{(2t+k)^2}{(2t+2k)^2} \right) \right\} .$$

We now use Theorem 3.19 to see that

$$f_t(k) C\left(\frac{t}{k}\right) t^{-1/2} \sim \frac{(2t+2k)^{2t+2k}}{k^k (2t+k)^{2t+k}} .$$

The asymptotics in this case appear similar to those for the binomial coefficient $\binom{2t+2k}{k}$. As opposed to the situation with Dyck paths, one may check that $f_t(k)$ is not equal to a binomial coefficient.

8. Discussion of Other Methods.

8.1. GF-Sequence Methods. The benchmark work in the area of multivariate asymptotics is still the 1983 article of Bender and Richmond [7]. Their main result was a local central limit theorem [7, Corollary 2] with the exact same conclusion as Theorem 3.29. Their hypotheses are as follows:

(i) $a_{\mathbf{r}} \geq 0$.

(ii) F has an algebraic singularity of order $q \notin \{0, -1, -2, \ldots\}$ on the graph of a function $z_d = g(z_1, \ldots, z_{d-1})$.

(iii) F is analytic and bounded in a larger polydisk, if one excludes a neighborhood of $\mathrm{Im}\,(z_j) = 0$ for each j.

(iv) B is nonsingular.

Comparing this to the results presented in this paper, we find both methodological and phenomenological differences. One methodological difference is that they view a d-variate generating function F as a sequence $\{F_n\}$ of $(d-1)$-variate generating functions. Their main result on coefficients of F is derived as a corollary of a result [7, Theorem 2] on sequences satisfying $F_n \sim C_n g h^n$ for some appropriately smooth g and h. As we have remarked, this approach is natural for some but not all applications and leads to some asymmetry in the hypotheses and conclusions.

A more important methodological difference is that while we always work in the analytic category, Bender and Richmond use a blend of analytic and smooth techniques.[3] This manifests itself in the hypotheses: where we require meromorphicity in a slightly enlarged polydisk, they require that the function g be in C^3, that the residue $(1 - z/g)^q F$ be in C^0, and that F be analytic away from the real coordinate planes. While our hypotheses are stronger in this regard, we know of no applications where their assumptions hold without $(1 - z/g)^q F$ being analytic. Bender and Richmond gain generality by allowing q to be nonintegral. This is further exploited by Gao and Richmond, where the singularity is allowed to be algebraico-logarithmic [24, Corollary 3]. On the other hand, their methods entail estimates and therefore cannot handle cancellation, such as occurs when $a_{\mathbf{r}}$ have mixed signs and the dominant singularity lies beyond the domain of convergence of F (see sections 4.10 and 4.3).

Phenomenologically, there is a significant difference in general between our methods and those of Bender, Richmond, Gao, et al. Their results govern only the case where a local central limit theorem holds—indeed it seems they are interested mainly or only in this case. Other behaviors of interest, which have been analyzed in the literature by various means, include Airy-type limits (see section 6.3), polynomial growth (see section 4.11), and elliptic-type limits. Central limit behavior results from smooth points $\mathbf{z}(\mathbf{r})$ with nondegenerate quadratic approximations to $h_{\mathbf{r}}(\mathbf{z}(\mathbf{r}))$, while Airy-type limits result from degenerate quadratic approximations, polynomial growth or corrections result when $\mathbf{z}(\mathbf{r})$ is a multiple point, and elliptic-type limits result from bad points. By restricting our exposition to the simplest cases, we have stayed mainly within the smooth point case, where $\{a_{\mathbf{r}}\}$ obeys an LCLT and the methods of Bender et al. apply, in most cases, equally as well as the results in section 3.3. But the advantage of multivariate analytic methods is that they can in principle be applied to any case in which F is meromorphic or algebraic. Thus, in addition to the further

[3] Analyticity is used only once, when they rotate the quadratic form B.

generality covered in section 3.4 of this paper, the same general method has been used to produce Airy limit results [57, 50] and is being applied to algebraic generating functions and meromorphic functions with bad point singularities, the simplest of which are quadratic cones. While this work is not yet published, it appears that it will provide another proof of the elliptic limit results for tiling statistics on the Aztec diamond [13] that is capable of generalizing to any quadratic cone singularity. This would prove similar behavior in two cases where such behavior is only conjectured, namely, cube groves [59] and quantum random walks (Chris Moore, personal communication), as well as unifying these results with the analyses of the coefficients of the generating function $1/(1-x-y-z+4xyz)$ that arise in the study of super ballot numbers [29] and Laguerre polynomials [30].

8.2. The Diagonal Method. There is a third method for obtaining multivariate asymptotics that deserves mention, namely, the so-called *diagonal method*. This derives a univariate generating function $f(z) = \sum_n a_{n,n} z^n$ for the main diagonal of a bivariate generating function $F(x, y) = \sum_{rs} a_{rs} x^r y^s$. The asymptotics of $a_{n,n}$ may then be read off by standard univariate means from the function f. The method may be adapted to compute a generating function for the coefficients $a_{np,nq}$ along any line of rational slope. This method, long known in various literatures, entered the combinatorics literature in [23, 37]; our exposition is taken from [67, section 6.3].

While this elegant method produces an actual generating function, which is more informative than the diagonal asymptotics, its scope is quite limited. First, while asymptotics in any rational direction may be obtained, the complexity of the computation of $a_{np,nq}$ increases with p and q. Thus there is no continuity of complexity, and no way to obtain uniform asymptotics or asymptotics in irrational directions. Second, the result is strictly bivariate. Third, even when the diagonal method may be applied, the computation is typically very unwieldy.

As an example, consider the generating function

$$F(x, y) = \sum_{m,n} a_{mn} x^m y^n = \frac{1 + xy + x^2y^2}{1 - x - y + xy - x^2y^2},$$

which [54] enumerates binary words without zigzags (a zigzag is defined to be a subword 010 or 101—the terminology comes from the usual correspondence of such words with Dyck paths, where $0, 1$, respectively, correspond to the steps $(1, 1), (1, -1)$). Here m, n, respectively, denote the number of 0's and 1's in the word. The main diagonal enumerates zigzag-free words with an equal number of 0's and 1's. The solutions to $H = 0, xH_x = yH_y$ are given by $x = y = 1/\phi$, $\phi = (1 + \sqrt{5})/2$, and $x = y = (1 + \sqrt{3}i)/2$. Thus contrib is a singleton $\{(1/\phi, 1/\phi)\}$ and the first order asymptotic is readily computed to be

$$a_{nn} \sim \phi^{2n} \frac{2}{\sqrt{n\pi\sqrt{5}}}.$$

The computation for any other diagonal is analogous, with the same amount of computational effort, and the asymptotics are uniform over any compact subset of directions keeping away from the coordinate axes.

To obtain the same result via the diagonal method requires the following steps. For each fixed t near 0, we compute the integral

$$D(z) := \sum_n a_{nn} z^n = \frac{1}{2\pi i} \int_{\mathcal{C}_t} F(z, t/z) \frac{dz}{z} = \frac{1}{2\pi i} \int_{\mathcal{C}_t} \frac{1 + t + t^2}{-z^2 + (1 + t - t^2)z - t},$$

where the contour is a circle that encloses all the poles of $F(z,t/z)/z$ satisfying $z(t) \to 0$ as $t \to 0$. Since $F(z,0)/z$ has a single simple pole at $z = 0$, the same is true of $F(z,t/z)/z$ for sufficiently small t. In this simple example we can explicitly solve for the pole $z(t)$ and compute its residue so that we obtain the result

$$\sum_n a_{nn} z^n = \sqrt{\frac{1+z+z^2}{1-3z+z^2}}\,.$$

In other cases, after some manipulation we obtain $\sum_n a_{nn} z^n$ implicitly as the solution of an algebraic equation. We are then faced with the problem of extracting asymptotics, which can probably be done using univariate techniques. However, we have left the realm of meromorphic series and this can complicate matters. In the example above, the branching occurs outside the domain of convergence, and the asymptotics are controlled by the dominant pole at ϕ^{-2} (the minimal zero of the denominator). Thus one obtains the same asymptotic as above, after some effort.

If we want to repeat this computation with $\sum_n a_{pn,qn} z^n$, we are required to find all small poles of the function $F(z^q, t/z^p)$. It is unlikely that these may be found explicitly, which complicates the task of finding which ones go to zero and computing the residues there.

Finally, another serious problem faced by the diagonal method is that while the diagonal of a rational series in $d = 2$ variables is always algebraic, a fact which can itself be proved by the diagonal method, in $d \geq 3$ variables, the diagonal of a rational series must be D-finite but may not be algebraic [48]. Thus a description of the diagonal generating function is more challenging. For example, consider the generating function $F(x,y,z) = (1-x-y-z)^{-1} = \sum a_{rst} x^r y^s z^t$, whose diagonal coefficient $a_{n,n,n}$ is the multinomial coefficient $\binom{3n}{n,n,n}$. This is known not to be algebraic, since its asymptotic leading term $C\alpha^n n^{-1}$ is not of the right form for an algebraic function. It is completely routine to derive this asymptotic using the methods of the present article, but any method that relies on an exact description of the diagonal will clearly require substantial extra work.

Acknowledgments. Many thanks to Manuel Lladser and Alex Raichev for a number of comments on preliminary drafts.

REFERENCES

[1] I. A. Aĭzenberg and A. P. Yuzhakov, *Integral Representations and Residues in Multidimensional Complex Analysis*, Transl. Math. Monogr. 58, AMS, Providence, RI, 1983.
[2] G. E. Andrews and K. Eriksson, *Integer Partitions*, Cambridge University Press, Cambridge, UK, 2004.
[3] C. Banderier, P. Flajolet, G. Schaeffer, and M. Soria, *Random maps, coalescing saddles, singularity analysis, and Airy phenomena*, Random Structures Algorithms, 19 (2001), pp. 194–246.
[4] Y. Baryshnikov and R. Pemantle, *Generating functioning with poles on a hyperplane arrangement*, in preparation.
[5] E. A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, J. Combin. Theory Ser. A, 15 (1973), pp. 91–111.
[6] E. A. Bender, *Asymptotic methods in enumeration*, SIAM Rev., 16 (1974), pp. 485–515.
[7] E. A. Bender and L. B. Richmond, *Central and local limit theorems applied to asymptotic enumeration.* II. *Multivariate generating functions*, J. Combin. Theory Ser. A, 34 (1983), pp. 255–265.
[8] E. A. Bender and L. B. Richmond, *Multivariate asymptotics for products of large powers with applications to Lagrange inversion*, Electron. J. Combin., 6 (1999), article 8 (electronic).

[9] E. A. Bender, L. B. Richmond, and S. G. Williamson, *Central and local limit theorems applied to asymptotic enumeration.* III. *Matrix recursions*, J. Combin. Theory Ser. A, 35 (1983), pp. 263–278.
[10] A. Bertozzi and J. McKenna, *Multidimensional residues, generating functions, and their application to queueing networks*, SIAM Rev., 35 (1993), pp. 239–268.
[11] M. Bousquet-Mélou and M. Petkovšek, *Linear recurrences with constant coefficients: The multivariate case*, Discrete Math., 225 (2000), pp. 51–75.
[12] F. Chung, R. Graham, J. Morrison, and A. Odlyzko, *Pebbling a chessboard*, Amer. Math. Monthly, 102 (1995), pp. 113–123.
[13] H. Cohn, N. Elkies, and J. Propp, *Local statistics for random domino tilings of the Aztec diamond*, Duke Math. J., 85 (1996), pp. 117–166.
[14] L. Comtet, *Advanced Combinatorics*, D. Reidel Publishing, Dordrecht, The Netherlands, 1974.
[15] D. A. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*, 2nd ed., Grad. Texts Math. 185, Springer-Verlag, New York, 2005.
[16] J. A. De Loera and B. Sturmfels, *Algebraic unimodular counting*, Math. Program., 96 (2003), pp. 183–203.
[17] E. Deutsch and L. Shapiro, *A survey of the Fine numbers*, Discrete Math., 241 (2001), pp. 241–265.
[18] M. Drmota, *A bivariate asymptotic expansion of coefficients of powers of generating functions*, European J. Combin., 15 (1994), pp. 139–152.
[19] R. Durrett, *Probability: Theory and Examples*, 3rd ed., Duxbury Press, Belmont, CA, 2004.
[20] P. Flajolet and A. Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math., 3 (1990), pp. 216–240.
[21] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, in preparation. Version 0.7 (2006) downloaded from http://algo.inria.fr/flajolet/Publications/books.html.
[22] A. Flaxman, A. W. Harrow, and G. B. Sorkin, *Strings with maximally many distinct subsequences and substrings*, Electron. J. Combin., 11 (2004), article 8 (electronic).
[23] H. Furstenberg, *Algebraic functions over finite fields*, J. Algebra, 7 (1967), pp. 271–277.
[24] Z. Gao and L. B. Richmond, *Central and local limit theorems applied to asymptotic enumeration.* IV. *Multivariate generating functions*, J. Comput. Appl. Math., 41 (1992), pp. 177–186.
[25] Z. Gao and N. C. Wormald, *The size of the largest components in random planar maps*, SIAM J. Discrete Math., 12 (1999), pp. 217–228.
[26] Z. Gao and N. C. Wormald, *The distribution of the maximum vertex degree in random planar maps*, J. Combin. Theory Ser. A, 89 (2000), pp. 201–230.
[27] D. Gardy, *Some results on the asymptotic behaviour of coefficients of large powers of functions*, Discrete Math., 139 (1995), pp. 189–217.
[28] I. M. Gel'fand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants*, Math. Theory Appl., Birkhäuser Boston, Boston, MA, 1994.
[29] I. M. Gessel, *Super ballot numbers*, J. Symbolic Comput., 14 (1992), pp. 179–194.
[30] J. Gillis, B. Reznick, and D. Zeilberger, *On elementary methods in positivity theory*, SIAM J. Math. Anal., 14 (1983), pp. 396–398.
[31] M. Goresky and R. MacPherson, *Stratified Morse Theory*, Ergeb. Math. Grenzgeb. (3) 14, Springer-Verlag, Berlin, 1988.
[32] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Dover, Mineola, NY, 2004.
[33] J. R. Griggs, P. Hanlon, A. M. Odlyzko, and M. S. Waterman, *On the number of alignments of k sequences*, Graphs Combin., 6 (1990), pp. 133–146.
[34] J. R. Griggs, P. J. Hanlon, and M. S. Waterman, *Sequence alignments with matched sections*, SIAM J. Algebraic Discrete Methods, 7 (1986), pp. 604–608.
[35] L. J. Guibas and A. M. Odlyzko, *String overlaps, pattern matching, and nontransitive games*, J. Combin. Theory Ser. A, 30 (1981), pp. 183–208.
[36] G. Hardy and S. Ramanujan, *Asymptotic formulae for the distribution of integers of various types*, Proc. London Math. Soc., 16 (1917), pp. 112–132.
[37] M. L. J. Hautus and D. A. Klarner, *The diagonal of a double power series*, Duke Math. J., 38 (1971), pp. 229–235.
[38] W. K. Hayman, *A generalisation of Stirling's formula*, J. Reine Angew. Math., 196 (1956), pp. 67–95.
[39] P. Henrici, *Applied and Computational Complex Analysis. Vol.* 2, Wiley Classics Library, John Wiley, New York, 1991.
[40] L. Hörmander, *An Introduction to Complex Analysis in Several Variables*, 3rd ed., North-Holland Math. Library 7, North-Holland, Amsterdam, The Netherlands, 1990.

[41] H.-K. Hwang, *Large deviations for combinatorial distributions.* I. *Central limit theorems*, Ann. Appl. Probab., 6 (1996), pp. 297–319.

[42] H.-K. Hwang, *Large deviations of combinatorial distributions.* II. *Local limit theorems*, Ann. Appl. Probab., 8 (1998), pp. 163–181.

[43] Y. Kogan, *Asymptotic expansions for large closed and loss queueing networks*, Math. Probl. Eng., 8 (2002), pp. 323–348.

[44] M. Larsen and R. Lyons, *Coalescing particles on an interval*, J. Theoret. Probab., 12 (1999), pp. 201–205.

[45] J. Leray, *Le calcul différentiel et intégral sur une variété analytique complexe. (Problème de Cauchy.* III*)*, Bull. Soc. Math. France, 87 (1959), pp. 81–180.

[46] S.-Y. R. Li, *A martingale approach to the study of occurrence of sequence patterns in repeated experiments*, Ann. Probab., 8 (1980), pp. 1171–1176.

[47] B. Lichtin, *The asymptotics of a lattice point problem associated to a finite number of polynomials.* I, Duke Math. J., 63 (1991), pp. 139–192.

[48] L. Lipshitz, *The diagonal of a D-finite power series is D-finite*, J. Algebra, 113 (1988), pp. 373–378.

[49] M. Lladser, *Uniform formulae for coefficients of meromorphic functions in two variables. Part* I, SIAM J. Discrete Math., 20 (2006), pp. 811–828.

[50] M. Lladser, *Asymptotic Enumeration via Singularity Analysis*, Ph.D. thesis, The Ohio State University, 2003.

[51] M. E. Lladser, P. Potočnik, J. Šiagiová, J. Širáň, and M. C. Wilson, *The diameter of random Cayley digraphs of given degree*, Random Structures Algorithms, submitted (2006); available online from http://arriv.org.labs/0706.3539.

[52] A. Meir and J. W. Moon, *The asymptotic behaviour of coefficients of powers of certain generating functions*, European J. Combin., 11 (1990), pp. 581–587.

[53] D. Mumford, *Algebraic Geometry.* I, Classics Math., Springer-Verlag, Berlin, 1995.

[54] E. Munarini and N. Zagaglia Salvi, *Binary strings without zigzags*, Sém. Lothar. Combin., 49 (2002/2004), article B49h (electronic); http://www.mat.univie.at/nsic/wpapers/s49/html.

[55] A. M. Odlyzko, *Asymptotic enumeration methods*, in Handbook of Combinatorics, Vols. 1, 2, Elsevier, Amsterdam, The Netherlands, 1995, pp. 1063–1229.

[56] R. Pemantle, *Generating functions with high-order poles are nearly polynomial*, in Mathematics and Computer Science (Versailles, 2000), Trends Math., Birkhäuser, Basel, 2000, pp. 305–321.

[57] R. Pemantle and M. C. Wilson, *Asymptotics of multivariate sequences.* I. *Smooth points of the singular variety*, J. Combin. Theory Ser. A, 97 (2002), pp. 129–161.

[58] R. Pemantle and M. C. Wilson, *Asymptotics of multivariate sequences.* II. *Multiple points of the singular variety*, Combin. Probab. Comput., 13 (2004), pp. 735–761.

[59] T. K. Petersen and D. Speyer, *An arctic circle theorem for groves*, J. Combin. Theory Ser. A, 111 (2005), pp. 137–164.

[60] G. Pólya, *On the number of certain lattice polygons*, J. Combin. Theory, 6 (1969), pp. 102–105.

[61] M. Régnier and F. Tahi, *Generating functions in computational biology*, J. Iranian Statist. Soc., to appear.

[62] J. Riordan, *Combinatorial Identities*, John Wiley, New York, 1968.

[63] D. G. Rogers, *Pascal triangles, Catalan numbers and renewal arrays*, Discrete Math., 22 (1978), pp. 301–310.

[64] L. W. Shapiro, S. Getu, W. J. Woan, and L. C. Woodson, *The Riordan group*, Discrete Appl. Math., 34 (1991), pp. 229–239.

[65] R. Sprugnoli, *Riordan arrays and combinatorial sums*, Discrete Math., 132 (1994), pp. 267–290.

[66] R. P. Stanley, *Enumerative Combinatorics. Vol.* 1, Cambridge Stud. Adv. Math. 49, Cambridge University Press, Cambridge, UK, 1997.

[67] R. P. Stanley, *Enumerative Combinatorics. Vol.* 2, Cambridge Stud. Adv. Math. 62, Cambridge University Press, Cambridge, UK, 1999.

[68] B. Sturmfels, *What is . . . a Gröbner basis?*, Notices Amer. Math. Soc., 52 (2005), pp. 1199–1200.

[69] J. van der Hoeven and R. Pemantle, *Automatic computation of asymptotics of coefficients of bivariate rational generating functions*, in preparation.

[70] M. S. Waterman, *Applications of combinatorics to molecular biology*, in Handbook of Combinatorics, Vols. 1, 2, Elsevier, Amsterdam, The Netherlands, 1995, pp. 1983–2001.

[71] H. S. Wilf, *generatingfunctionology*, 2nd ed., Academic Press, Boston, MA, 1994.

[72] M. C. Wilson, *Asymptotics for generalized Riordan arrays*, in Proceedings of the 2005 International Conference on Analysis of Algorithms, Discrete Math. Theor. Comput. Sci. Proc., AD, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2005, pp. 323–333 (electronic).