Chapter 2

Generating functions in one and several variables

This chapter gives a crash course on generating functions and enumeration. For a more lengthy introduction, I recommend [Wil94]. Chapter 14 of [vLW92] also provides a fairly concise but readable treatment. Proofs for facts about formal power series may be found in [Sta97, Section 1.1]. A comprehensive treatment of the relation between power series operations and corresponding combinatorial constructions on finite sets is the encyclopedic reference [GJ83]. Chapters I–III of the preprint [FS08] contain a very nice treatment as well.

Throughout the lecture notes, but particularly in this chapter, the notation $[n]$ will denote the set $\{1, \ldots, n\}$.

2.1 Formal power series

Let $\mathbb{C}[[z_1, \ldots, z_d]]$ denote the ring of **formal power series** in the variables $z_1, \ldots, z_d$. Elements of $\mathbb{C}[[z_1, \ldots, z_d]]$ are parameterized by collections $\{f_\mathbf{r} : \mathbf{r} \in (\mathbb{Z}^+)\}$ of complex numbers via the correspondence $\{f_\mathbf{r}\} \mapsto \sum_\mathbf{r} f_\mathbf{r} \mathbf{z}^\mathbf{r}$. Addition is defined by $(f + g)_\mathbf{r} = f_\mathbf{r} + g_\mathbf{r}$ and multiplication is defined by convolution: $(f \cdot g)_\mathbf{r} = \sum_\mathbf{s} f_\mathbf{s} g_{\mathbf{r}-\mathbf{s}}$. The sum in this convolution is always finite, so there is no question of convergence. In other words, any array $\{f_\mathbf{r} : \mathbf{r} \in (\mathbb{Z}^+)^d\}$ has a well-defined generating function in the formal power series ring. A common notation for $f_\mathbf{r}$ is $[\mathbf{z}^\mathbf{r}]\, f$, read as "the $\mathbf{z}^\mathbf{r}$-coefficient of f".

The additive identify in $\mathbb{C}[[z_1, \ldots, z_d]]$ is the zero series and the multiplicative identity is the function that maps the vector $\mathbf{0}$ to 1 and everything else to zero. It is an easy exercise to see that f has a multiplicative inverse if and only if $f_\mathbf{0} \neq 0$. Thus $\mathbb{C}[[z_1, \ldots, z_d]]$ is a **local ring**, meaning there is a unique maximal ideal, $\mathfrak{m}$, the set of non-units. Local rings come equipped with a notion of convergence, namely $f_n \to f$ if and only if $f_n - f$ is eventually in $\mathfrak{m}^k$ for every k. An easier way to say this is that for all $\mathbf{r}$ there is an $N(\mathbf{r})$ such that $(f_n)_\mathbf{r} = f_\mathbf{r}$ for $n \geq N(\mathbf{r})$.

An open **polydisk** centered at a complex point $\mathbf{z} \in \mathbb{C}^d$ is the set of $\mathbf{y} \in \mathbb{C}^d$ such that $|y_j - z_j| < b_j$ for all $1 \leq j \leq d$, where $b_j > 0$ are specified constants. Let $\mathcal{N}$ be an open polydisk centered at the origin in $\mathbb{C}^d$, that is a set $\{\mathbf{z} : |z_i| < t_i, 1 \leq i \leq d\}$. Suppose that $f, g \in \mathbb{C}[[z_1, \ldots, z_d]]$ are absolutely convergent on $\mathcal{N}$, that is, $\sum_\mathbf{r} |f_\mathbf{r}||w_1|^{r_1} \cdots |w_d|^{r_d} < \infty$ when all $|w_i| < t_i$, and similarly for g. Then $f + g$ and $f \cdot g$ are absolutely convergent on $\mathcal{N}$ as well and the sum and product in the ring of formal power series is the same as in the ring of analytic functions in $\mathcal{N}$. Since the intersection of neighborhoods of the origin is a neighborhood of the origin, the subset of $\mathbb{C}[[z_1, \ldots, z_d]]$ of series that converge in some neighborhood of the origin is a subring. This is called the ring of germs of analytic functions, and is not all of $\mathbb{C}[[z_1, \ldots, z_d]]$. That is, there are some formal power series that fail to converge anywhere (except at the origin) and for these it will not work to apply analytic methods. One can however make a generating function by letting $F(\mathbf{x}) = \sum_\mathbf{r} f_\mathbf{r} \mathbf{z}^\mathbf{r} / g(\mathbf{r})$ for a judiciously chosen g. A good choice is often to let $g(\mathbf{r})$ be a product of some or all of the quantities $r_i!$; a generating function normalized by factorials is called an **exponential generating function**. Not only may this normalization cause the power series to converge, but the behavior of exponential generating functions under convolution has an important combinatorial interpretation. Several examples of this are given in Section 2.5.

One can go backwards as well. If $f = g/h$ where g and h are convergent on a neighborhood of the origin and $h \notin \mathfrak{m}$, then f is analytic on a neighborhood of the origin, in fact it is analytic where g and h are and where h is nonzero, and its Taylor series is equal to g/h in $\mathbb{C}[[z_1, \ldots, z_d]]$. Similarly, one may define formal differentiation by

$$\frac{\partial}{\partial z_j} F = \sum_\mathbf{r} r_j a_\mathbf{r} \mathbf{z}^{\mathbf{r} - \delta_j}$$

and this will agree with analytic differentiation on the domain of convergence of F.

The interior $\mathcal{D}$ of the domain on which the formal power series F converges is the union of open polydisks. In particular it is the union of tori, and is hence characterized by its intersection $\mathcal{D}_\mathbb{R}$ with $\mathbb{R}^d$. The set $\mathcal{D}$ is in fact pseudoconvex, meaning that the set $\operatorname{Log} \mathcal{D}$ defined by $(x_1, \ldots, x_d) \in \operatorname{Log} \mathcal{D}$ iff $(e^{x_1}, \ldots, e^{x_d}) \in \mathcal{D}$ is a convex order ideal[1]. See Hörmander (1990, Section 2.5) for these and other basic facts about functions of several complex variables.

Just as we use r, s and t for r_1, r_2 and r_3, so as to make examples more readable we will use x, y and z for z_1, z_2 and z_3.

2.2 Rational operations on generating functions

A d-variate **combinatorial class** is a set $\mathcal{A}$ which is the disjoint union of finite sets $\{\mathcal{A}_\mathbf{r} : \mathbf{r} \in (\mathbb{Z}^+)^d\}$ in some natural way. In this section, $F = \sum_\mathbf{r} a_\mathbf{r} \mathbf{z}^\mathbf{r}$ will "generate" a combinatorial class $\mathcal{A}$, that is,

[1] An order ideal is a set closed under $\leq$ in the coordinatewise partial order on $\mathbb{R}^d$.

$|\mathcal{A}_{\mathbf{r}}| = a_{\mathbf{r}}$ for all $\mathbf{r}$. We also say that F "counts $\mathcal{A}$ by ϕ" where ϕ is the map taking $x \in \mathcal{A}$ to the $\mathbf{r}$ for which $x \in \mathcal{A}_{\mathbf{r}}$. Arithmetical operations in the ring of formal power series were defined so as to correspond to existing operations on analytic power series. It is instructive to find interpretations for these operations on the combinatorial level. Here follows a list of set-theoretic interpretations for rational operations. The combinatorial wealth of these interpretations explains why there are so many rational generating functions in combinatorics.

Equality: bijection

It goes almost without saying that equality between two generating functions F and G corresponds to bijective correspondence between the classes they generate: $|\mathcal{A}_{\mathbf{r}}| = |\mathcal{B}_{\mathbf{r}}|$ for all $\mathbf{r}$.

Multiplication by z_j: re-indexing

In the univariate case, the function $zF(z)$ generates the sequence $0, a_0, a_1, a_2, \ldots$. Similarly, in the multivariate case, $z_j F(\mathbf{z})$ generates $\{b_{\mathbf{r}}\}$ where $b_{\mathbf{r}} = a_{\mathbf{r}-\delta_j}$, which is defined to be zero if any coordinate is negative.

Re-indexing the other direction is more complicated. In the univariate case, the sequence $a_1, a_2, \ldots$ is generated by the function $(f - f(0))/z$. In the multivariate case, the sequence $\{a_{\mathbf{r}+\delta_j}\}$ is generated by $(F - F(z_1, \ldots, z_{j-1}, 0, z_{j+1}, \ldots, z_d))/z_j$.

Sums: disjoint unions

If F generates a class $\mathcal{A}$ and G generates a class $\mathcal{B}$, then $F + G$ generates the class $\mathcal{C}$ where $\mathcal{C}_{\mathbf{r}}$ is the disjoint union of $\mathcal{A}_{\mathbf{r}}$ and $\mathcal{B}_{\mathbf{r}}$. The interpretations of equality, multiplication by z_j and sums on the combinatorial level are pretty simple, but already one may find examples that are not completely trivial.

Example 2.1 (binary sequences with no repeated 1's). Let $\mathcal{A}_n$ be the set of sequences of 0's and 1's of length n that do not begin with 1 and have no two consecutive 1's. Each such sequence ends either in 0 or in 01. The sequence that remains can be any sequence in $\mathcal{A}_{n-1}$ or $\mathcal{A}_{n-2}$ respectively. Thus, stripping off the last one or two symbols respectively yields a bijective correspondence between $\mathcal{A}_n$ and the disjoint union $\mathcal{A}_{n-1} \cup \mathcal{A}_{n-2}$. At the generating function level, we see that $F(z) = zF(z) + z^2 F(z)$ — well, almost! If we take $\mathcal{A}_n$ to be empty for $n < 0$, the correspondence still works for $n = 1$ but if fails for $n = 0$. Thus, actually,

$$zF(z) + z^2 F(z) = F(z) - 1\,.$$

Via operations in the formal power series ring we may rearrange to isolate the 1 and then divide by $1 - z - z^2$ to obtain

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

Example 2.2 (binomial coefficients). Let $\mathcal{A}_{r,s}$ be the set of colorings of the set $[r+s] := \{1,\ldots,r+s\}$ for which r elements are red and s are green. Decomposing according to the color of the last element, $\mathcal{A}_{r+s}$ is in bijective correspondence with the disjoint union of $\mathcal{A}_{r-1,s}$ and $\mathcal{A}_{r,s-1}$. This is a combinatorial interpretation of the identity $\binom{r+s}{r} = \binom{r+s-1}{r} + \binom{r+s-1}{r-1}$ and holds as long as $r+s>0$. It follows that

$$F(x,y) - 1 = xF(x,y) + yF(x,y)$$

and solving for F gives

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

Products: convolutions

If F generates the class $\mathcal{A}$ and G generates the class $\mathcal{B}$ then FG generates the class $\mathcal{C}$ defined by letting $\mathcal{C}_{\mathbf{r}}$ be the disjoint union of cartesian products $\mathcal{A}_{\mathbf{s}} \times \mathcal{B}_{\mathbf{r}-\mathbf{s}}$ over all $\mathbf{s} \leq \mathbf{r}$. This is the canonical definition of a product in any category of graded objects.

Students of probability theory will recognize it as a convolution. Suppose that F and G have nonnegative coefficients. Suppose furthermore that $F(\mathbf{1}) = G(\mathbf{1}) = 1$, where $\mathbf{1}$ is the d-vector of 1's (that is, the coefficients of each sum to 1). Then F is the **probability generating function** for a probability distribution on $(\mathbb{Z}^+)^d$ that gives mass $a_{\mathbf{r}}$ to the point $\mathbf{r}$, and G is likewise a probability generating function. The convolution FG generates the distribution of the sum of independent picks from the two probability distributions. Thus the study of sums of independent, identically distributed random variables taking values in $(\mathbb{Z}^+)^d$ is equivalent to the study of powers of such a generating function F. The laws of large numbers in probability theory may be derived via generating function analyses, while the central limit theorem is always proved essentially this way. In Chapter 10, versions of these laws are proved for coefficients of generating functions far more general than powers of probability generating functions.

A useful trick with products is as follows.

Example 2.3 (enumerating partial sums). Let $F(z)$ enumerate the class $\mathcal{A}$ and let $G(z) = 1/(1-z)$ enumerate a class $\mathcal{B}$ with $|\mathcal{B}_n| = 1$ for all n. Then FG enumerates the class $\mathcal{C}$ where $\mathcal{C}_n$ is the disjoint union of $\biguplus_{j=0}^n \mathcal{A}_j$. Consequently, the generating function for the partial sums $\sum_{j=0}^n a_j$ is $F(z)/(1-z)$.

The operation $1/(1-F)$: finite sequences

Let $\mathcal{B}$ be a combinatorial class and let $\mathcal{A}$ be the class of finite sequences of elements of $\mathcal{B}$, graded by total weight, meaning that the sequence $(x_1,\ldots,x_k)$ belongs to $\mathcal{A}_{\mathbf{r}}$ if $x_j \in \mathcal{B}_{\mathbf{s}^{(j)}}$ for $1 \leq j \leq k$ and $\sum_{j=1}^k \mathbf{s}^{(j)} = \mathbf{r}$. Then $\mathcal{A}$ is the disjoint union of the empty sequence, the class of singleton sequences, the class of sequences of length 2, and so forth, and summing the generating functions gives $F = 1+G+G^2+\cdots$. Provided that $\mathcal{B}$ has no elements of weight zero ($\mathcal{B}_0 = \emptyset$), this converges in the ring of formal power series and is equal to $1/(1-G)$. If G grows no faster than exponentially, then both sides of this equation converge analytically in a neighborhood of the origin, and are equal to $1/(1-G)$.

A simple example of this is to count the binary strings of Example 2.1 by the number of zeros and the number of ones, rather than by total length. Any such sequence may be uniquely decomposed into a finite sequence of the blocks 0 and 01. Letting these have weights $(1,0)$ and $(1,1)$ respectively, the class $\mathcal{B}$ of blocks has generating function $x + xy$. The generating function for $\mathcal{A}$ is therefore $F(x,y) = 1/(1-x-xy)$. We may collapse this to a univariate function by using weights 1 and 2 instead of $(1,0)$ and $(1,1)$, recovering the generating function $1/(1-z-z^2)$.

Example 2.4 (prefix codes). Let T be a finite rooted binary tree (every vertex has either 0 or 2 children) whose vertices are identified with finite sequences of 0's and 1's. Any sequence of 0's and 1's may be decomposed into blocks by repeatedly stripping off the initial segment that is a leaf of T. The decomposition is unique; it may end with a partial block, that is, an internal node of T.

Here is a derivation of the generating function F counting all binary sequences by their length and the number of blocks. Let $B(x)$ be the univariate generating function counting leaves of T by depth. The generating function for blocks by length and number of blocks is $yB(x)$ since each block has number of blocks equal to 1. The generating function for binary sequences with no incomplete blocks is therefore $1/(1-yB(x))$. Allowing incomplete blocks, each sequence uniquely decomposes into a maximal sequence of complete blocks, followed by a (possibly empty) incomplete block. Letting $C(x)$ count incomplete blocks by length, we see that $1 + y(C(x)-1)$ counts incomplete blocks by length and number of blocks, so we have finally,

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

Lattice paths yield a large and well studied class of examples.

Example 2.5. Let E be a finite subset of $(\mathbb{Z}^+)^d$ not containing $\mathbf{0}$ and let $\mathcal{A}$ be the class of finite sequences $(\mathbf{0} = x_0, x_1, \ldots, x_k)$ of elements of $(\mathbb{Z}^+)^d$ with $x_j - x_{j-1} \in E$ for $1 \le j \le k$. We call these **paths with steps in** E.

Let $B(\mathbf{z}) = \sum_{\mathbf{r}\in E} \mathbf{z}^{\mathbf{r}}$ generate E by step size. Then $1/(1-B(\mathbf{z}))$ counts paths with steps in E by ending location. This includes examples we have already seen. Multinomial coefficients count paths ending at $\mathbf{r}$ with steps in the standard basis directions $e_1, \ldots, e_d$; the generating function $1/(1-\sum_{j=1}^d z_j)$ follows from the generating function $\sum_{j=1}^d z_j$ for E.

Example 2.6 (Delannoy numbers). Let $\mathcal{A}_{\mathbf{r}}$ be the lattice paths from the origin to $\mathbf{r}$ in $\mathbb{Z}^2$ using only steps that go North, East or Northeast to the next lattice point. The numbers $a_{\mathbf{r}} := |\mathcal{A}_{\mathbf{r}}|$ are called **Delannoy numbers** [Com74, Exercise I.21]. The generating function $x+y+xy$ for E leads to the Delannoy generating function

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

One final example comes from the paper [CLP04].

Example 2.7 (no gaps of size 2). Let $\mathcal{B}_n$ be the class of subsets of $[n]$ where no two consecutive members are absent. It is easy to count $\mathcal{B}_n$ by mapping bijectively to Example 2.1. However, in [CLP04], an estimate was required on the number of such sets that were mapped into other such sets by a random permutation. To compute this (actually to compute the second moment of this random variable) it sufficed to count the pairs $(S,T) \in \mathcal{B}_n^2$ by $n, |S|, |T|$ and $|S \cap T|$.

A 4-variable generating function $F(x,y,z,w)$ may be derived by investigating what may happen between consecutive elements of $S \cap T$. Identify $(S,T) \in \mathcal{B}_n^2$ with a sequence α in the set $\{(0,0),(0,1),(1,0),(1,1)\}^n$ where a $(1,1)$ in position j denotes an element of $S \cap T$, a $(1,0)$ denotes an element of $S \setminus T$, and so forth. If j and $j+r$ are positions of consecutive occurrences of $(1,1)$, then the possibilities for the string $\alpha_{j+1} \cdots \alpha_{j+r}$ are as follows:

(i) $\alpha_{j+1} = (1,1)$: the only possibility is $r = 1$.

(ii) $\alpha_{j+1} = (0,0)$: the only possibility is $r = 2$ and $\alpha = ((0,0),(1,1))$.

(iii) $\alpha = (1,0)$: then $r \geq 2$ may be arbitrary and α alternates between $(1,0)$ and $(0,1)$ until the final $(1,1)$.

(iv) $\alpha = (0,1)$: then $r \geq 2$ may be arbitrary and α alternates between $(0,1)$ and $(1,0)$ until the final $(1,1)$.

In the first case, the generating function $G_1(x,y,z,w)$ for blocks by the four weights is just $xyzw$. In the second case, $G_2(x,y,z,w) = x^2yzw$. In the third case, one may write the block as either $((1,0))$ or $((1,0),(0,1))$, followed by zero or more alternations of length two; decomposing this way shows the generating function to be $G_3(x,y,z,w) = xyzw\dfrac{xy+x^2yz}{1-x^2yz}$. Similarly, we see that $G_4(x,y,z,w) = xyzw\dfrac{xz+x^2yz}{1-x^2yz}$. Summing these gives a block generating function of

$$G(x,y,z,w) = xyzw\frac{(1+x)(1-x^2yz)+xy+xz+2x^2yz}{1-x^2yz}\,.$$

Finally, we use the $1/(1-G)$ formula. Stringing together blocks of the four types gives all legal sequences of any length that end in $(1,1)$; thus this class has generating function $1/(1-G)$. These correspond to pairs $(S,T) \in \mathcal{B}_n^2$ with $n \in S \cap T$, and are in bijective correspondence (via deletion of the element n) to all pairs in $\mathcal{B}_{n-1}^2$ except that when $n = 0$ it is not possible to delete n. The bijection reduces the weight of each (S,T) by $(1,1,1,1)$. Thus,

$$F(x,y,z,w) = \frac{\frac{1}{1-G(x,y,z,w)} - 1}{xyzw} = \frac{(1+x)(1-x^2yz)+xy+xz+2x^2yz}{1-x^2yz-xyzw[(1+x)(1-x^2yz)+xy+xz+2x^2yz]}\,.$$

Transfer matrices: restricted transitions

Suppose we want to count words (that is, finite sequences) in an alphabet V but only some consecutive pairs are allowed. Let E be the set of allowed pairs. Allowed words of length n are equivalent to paths of length n in the directed graph (V, E). To count these by length, let $\mathbf{M}$ be the incidence matrix of (V, E), that is, the square matrix indexed by V, with $M_{vw} = 1$ if $(v, w) \in E$ and $M_{vw} = 0$ otherwise. The number of allowed paths of length n from v to w is $(\mathbf{M}^n)_{vw}$. If we wish to count paths by length, we must sum $(z\mathbf{M})^n$ over n. Thus, the generating function counting finite paths from v to w by their length is

$$F(z) = \sum_{n=0}^{\infty} ((z\mathbf{M})^n)_{vw} = \left[(\mathbf{I} - z\mathbf{M})^{-1}\right]_{vw} .$$

This formula is quite versatile. To count all allowed paths by length, we may sum in v and w; a convenient way to notate this is $\text{trace}\left((\mathbf{I} - z\mathbf{M})^{-1}\mathbf{J}\right)$, where J is the $|V| \times |V|$ square matrix of 1's. Alternatively, we may count by features other than length. The most general way to count is by the number of each type of transition: enumerate $E = \{e_1, \dots, e_k\}$ and let $\tilde{M}_{vw} = z_k$ if $e_k = (v, w)$ and $\tilde{M}_{vw} = 0$ if $(v, w) \notin E$; then $[(\mathbf{I} - \tilde{\mathbf{M}})^{-1}]_{vw}$ counts paths from v to w by transitions and $\text{trace}\left((\mathbf{I} - \tilde{\mathbf{M}})^{-1}\mathbf{J}\right)$ counts all paths by transitions.

Example 2.8 (binary strings revisited). The transfer matrix method may be used to count the paths of Example 2.1. Let $V = \{0, 1\}$ and $E = \{(0, 0), (0, 1), (1, 0)\}$ contain all directed edges except $(1, 1)$. Then

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

so

$$\mathbf{Q} := (\mathbf{I} - z\mathbf{M})^{-1} = \frac{1}{1 - z - z^2} \begin{bmatrix} 1 & z \\ z & 1 - z \end{bmatrix}$$

The paths from 0 to 0 having n transitions, $n \geq 0$, are in one to one correspondence, via stripping off the last 0, to the words in Example 2.1 of length n. Thus the generating function is the $(0, 0)$-entry of $\mathbf{Q}$, namely $1/(1 - z - z^2)$.

Composition: block substitution

Let F be a d-variate generating function and $G_1, \dots, G_d$ be d generating functions in any number of variables, all with vanishing constant terms. One may define the formal composition $F \circ (G_1, \dots, G_d)$ as a limit in the formal power series ring:

$$F \circ (G_1, \dots, G_d) := \lim_{n \to \infty} \sum_{|\mathbf{r}| \leq n} a_{\mathbf{r}} \mathbf{G}^{\mathbf{r}} . \tag{2.1}$$

The degree of any monomial in $\mathbf{G}^{\mathbf{r}} := G_1^{r_1} \cdots G_d^{r_d}$ is at least $|\mathbf{r}| := \sum_{j=1}^{d} r_j$ by the assumption that $G_j(\mathbf{0}) = 0$ for all j, hence the $\mathbf{z}^{\mathbf{r}}$-coefficient of the sum does not change once $n > |\mathbf{r}|$ and the limit exists in the formal power series ring. Even if some $G_j(0) \neq 0$, it may still happen that the sum converges in the ring of analytic functions, meaning that the infinitely many contributions to all coefficients are absolutely summable.

A slightly unwieldy abstract combinatorial interpretation of this is given in Sections 2.2.20–2.2.22 of [GJ83]. Let $\mathcal{A}, \mathcal{B}_1, \ldots, \mathcal{B}_d$ be the classes generated respectively by $F, G_1, \ldots, G_d$. The class corresponding to $F \circ (G_1, \ldots, G_d)$ is obtained as a disjoint union over elements $x \in \mathcal{A}$ of d-tuples $(C_1, \ldots, C_d)$, where C_i is a sequence of length r_i of elements of $\mathcal{B}_i$. The weight of such a d-tuple is the sum of the weights of the C_i, which are in turn the sum of the weights of all r_i elements of $\mathcal{B}_i$. The following examples should clarify this.

Example 2.9 (queries). Queries from a data base have integer computation times associated with them. There are b_k queries of size k, $k \geq 1$. The protocol does not allow two long queries in a row, where long is defined as of size greater than some number M. How many query sequences are there of total time n?

The sequences of queries are bijectively equivalent to the composition $\mathcal{A} \circ (\mathcal{B}_1, \mathcal{B}_2)$, where $\mathcal{A}$ is the class from Example 2.8, counted by numbers of 0's and 1's, and $\mathcal{B}_1$ and $\mathcal{B}_2$ are respectively the short queries and the long queries, counted by computation time. Thus the function $F(G_1, G_2)$ counts queries by time, where $F(x, y) = 1/(1 - x - xy)$, $G_1(z) = \sum_{k=1}^{M} b_k z^k$ and $G_2(z) = \sum_{k>M} b_k z^k$.

The following example may seem a natural candidate for the transfer matrix method, but it is simpler to analyze it as from the viewpoint of compositions.

Example 2.10 (Smirnov words). Let $\mathcal{A}$ be the class of **Smirnov words**, that is words in the alphabet $[d]$ with no consecutive repetition of any symbol allowed. Of course $|\mathcal{A}_n| = d \cdot (d-1)^{n-1}$, but suppose we wish to count differently.

Let F count Smirnov words by number of occurrences of each symbol and let G count the class $\mathcal{B}$ of all words on the alphabet $[d]$, also by number of occurrences of each symbol. Starting with $x \in \mathcal{A}$ and substituting an arbitrary nonzero string of the symbol j for each occurrence of j in x produces each element of $\mathcal{B}$ in a unique way. The generating function for a nonempty string of j's is $\dfrac{z_j}{1 - z_j}$, so we have

$$G(\mathbf{z}) = F\left(\frac{z_1}{1 - z_1}, \ldots, \frac{z_d}{1 - z_d}\right).$$

Solve for F by setting $y_j = z_j/(1 + z_j)$ to obtain

$$F(\mathbf{y}) = G\left(\frac{y_1}{1 + y_1}, \ldots, \frac{y_d}{1 + y_d}\right)$$

and use $G(\mathbf{z}) = 1/(1 - \sum_{j=1}^{d} z_j)$ to get

$$F(\mathbf{z}) = \frac{1}{1 - \sum_{j=1}^{d} \frac{z_j}{1+z_j}}.$$

One subject in probability theory, namely the study of branching processes, is almost always dealt with by means of generating functions.

Example 2.11 (Galton-Watson process). Let $f(z)$ be a probability generating function, that is $f(z) = \sum_{n=0}^{\infty} p_n z^n$ with $p_n \geq 0$ and $\sum_{n=0}^{\infty} p_n = 1$. A **branching process** with offspring distribution f is a random family tree with one progenitor in generation 0 and each individual in each generation having a random number of children; these numbers of children born to the individuals in a generation are independent and each is equal to n with probability p_n. The random number of individuals in generation n is denoted Z_n. What is the probability $p_{n,k}$ of $Z_n = k$?

We compute the probability generating function for Z_n inductively as follows. The probability generating function for Z_1 is just f. Suppose we know the probability generating function $g_n := \sum_k p_{n,k} z^k$ for Z_n. Interpret this as saying that there are a total mass of $p_{n,k}$ configurations with $Z_n = k$. In a configuration with $Z_n = k$, the next generation is composed of a sequence of k families, each independently having size j with probability p_j. The probability generating function for such a sequence is f^k, hence we see that $g_{n+1} = g_n \circ f$. Inductively then, $g_n = f^{(n)} := f \circ \cdots \circ f$, a total of n times. Observe that, unless $p_0 = 0$ (no extinction), this composition is not defined in the formal power series ring, but since all functions involved are convergent on the unit disk, the compositions are well defined analytically.

Example 2.12 (Branching random walk). Associate to each particle in a branching process a real number, which we interpret as the displacement in one dimension between its position and that of its parent. If these are independent of each other and of the branching, and are identically distributed, then one has the classical branching random walk. A question that has been asked several times in the literature, e.g., [Ald05, Kes78], beginning with a single particle, say at position 1, is whether there is any line of descent that remains to the right of the origin for all time.

To analyze this, modify the process so that X denotes the number of particles ever to hit the origin. Let us examine this in the simplest case, where the branching process is deterministic binary splitting ($p_2 = 1$) and the displacement distribution is a random walk that moves one unit to the right with probability $p < 1/2$ and one unit to the left with probability $1 - p$. If we modify the process so that particles stop moving or reproducing when they hit the origin, then an infinite line of descent to the right of the origin is equivalent to infinitely many particles reaching the origin.

To analyze the process, therefore, we let X be the number of particles ever to hit the origin (still begin with a single particle at 1). Let ϕ be the probability generating function for X:

$$\phi(z) = \sum_{n=2}^{\infty} a_n z^n \quad ; \quad a_n := \mathbb{P}(X = n) .$$

If the initial condition is changed to a single particle at position 2, then the number of particles ever to reach the origin will have probability generating function $\phi \circ \phi$. To see this, apply the analysis of the previous example, noting that the number of particles ever to reach 1 before any ancestor has reached 1, together with their collections of descendants who ever reach 0, form two generations of a branching process with offspring distribution the same as X.

Each of the two children in the first generation is located at 0 with probability $1-p$ and at 2 with probability p, so the probability generating function for the contribution to X of each child is $(1-p)z + p\phi(\phi(z))$. The two contributions are independent so their sum is a convolution, whose probability generating function is therefore the square of this. Thus we have the identity

$$\phi(z) = [(1-p)z + p\phi(\phi(z))]^2 \,. \tag{2.2}$$

While this does not produce an explicit formula for ϕ, it will be shown in Example 3.10 below that there is enough information in (2.2) to derive asymptotics.

2.3 Algebraic generating functions

After rational functions, most people consider algebraic functions to be the next simplest class. These arise frequently in combinatorics as well. One reason, having to do with recursions satisfied by algebraic functions, will be taken up in the next section. Another reason is that when a combinatorial class solves a convolution equation, its generating function solves an algebraic equation. A famous univariate example of this is as follows.

Example 2.13 (binary trees and Catalan numbers). Let $\mathcal{A}$ be the class of finite, rooted, binary trees. This class is defined recursively as follows: the empty tree (no vertices) is in $\mathcal{A}$; every element of $\mathcal{A}$ with $n \geq 1$ vertices has a root which has a left and a right subtree; the possible ordered pairs (L, R) of left and right subtrees are just all ordered pairs of previously defined trees the cardinalities of which sum to $n-1$. Let $\mathcal{A}_n$ denote the subclass of binary trees with n vertices. The cardinality of $\mathcal{A}_n$ is the n^{th} Catalan number, usually denoted C_n; binary trees are one of dozens and dozens of classes counted by the Catalan numbers. A by no means exhaustive list of 66 of these is given in [Sta99, Problem 6.19].

The recursion implies that for $n \geq 1$, the set $\mathcal{A}_n$ is in bijection with the disjoint union $\mathcal{B}_n := \biguplus_{k=0}^{n-1} \mathcal{A}_k \times \mathcal{A}_{n-1-k}$. This is a re-indexed convolution of $\mathcal{A}$ with itself. At the level of generating functions, we see that the re-indexed convolution has generating function $zF(z)^2$. Taking into account what happens when $n = 0$ yields

$$F(z) - 1 = zF(z)^2 \,. \tag{2.3}$$

This ought to have a unique formal power series solution because any solution to this obeys the defining recursion and the initial condition for the class of binary trees. To solve (2.3) in the ring of formal power series, let us first find solutions in the ring of germs of analytic functions, since we know more operations there. The quadratic formula yields two solutions:

$$F(z) = \frac{1 \pm \sqrt{1-4z}}{2z} \,.$$

If either of these two functions is analytic near zero, then the series must have coefficients C_n because this is the unique solution to the convolution equation represented by $F = 1 + zF^2$. Because the denominator vanishes at zero, division can be valid only if the numerator also vanishes at zero. Choosing the negative root yields such a numerator, and one may check by rationalizing the denominator that the resulting function is analytic near zero:

$$F(z) = \frac{1 - \sqrt{1-4z}}{2z} = \frac{2}{1 + \sqrt{1-4z}}$$

which is analytic in the disk $|z| < 1/4$.

The kernel method

The rest of this section is devoted to the **kernel method**, one of the most prolific sources of algebraic generating functions. The kernel method is a means of producing a generating function for an array $\{a_{\mathbf{r}}\}$ satisfying a linear recurrence

$$a_{\mathbf{r}} = \sum_{\mathbf{s} \in E} c_{\mathbf{s}} a_{\mathbf{r}-\mathbf{s}} \tag{2.4}$$

for some constants $\{c_{\mathbf{s}} : \mathbf{s} \in E\}$, except when $\mathbf{r}$ is in the **boundary condition**, which will be made precise later. The set E is a finite subset of $\mathbb{Z}^d$ but not necessarily of $(\mathbb{Z}^+)^d$. Indeed, if $E \subseteq (\mathbb{Z}^+)^d$, then Example 2.5 generalizes easily to show that $F(\mathbf{z}) = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is rational. There is one further condition on E: its convex hull must not intersect the negative orthant $\{\mathbf{r} : \mathbf{r} \leq \mathbf{0}\}$. This ensures that the recursion is well founded (Lemma 2.15 below).

The kernel method is of interest to the present study because it often produces generating functions which, even though they are not rational, satisfy the meromorphicity assumptions that allow us to compute their asymptotics. It is shown in [BMP00] that the complexity of F increases with the number of coordinates in which points of E are allowed to take negative values. Just as allowing no negative coordinates in E causes F to be rational, it turns out that allowing only one negative coordinate in E causes F to be algebraic. This is shown in [BMP00], along with counterexamples when E contains points with two different negative coordinates. The remainder of this section draws heavily on [BMP00]. We begin, though, with an example.

Example 2.14 (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 immediately preceding the finish square, then 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 coordinate 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 $x^0 y^1$ coefficient corresponds is $2a_{0,0} - a_{0,-1} - a_{-1,1}$ which is equal to 2, not 0, because the recursion does not hold at $(0,0)$ (a_{00} is set equal to 1); the $y^0 x^j$ 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

$$Q(x,y)F(x,y) = 2y - h(x) \tag{2.5}$$

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 [LL99] 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 (2.5) gives

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

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

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

A general explanation of the kernel method

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 $\mathbf{p} \leq \mathbf{s}$ 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. We assume $\mathbf{p} \neq \mathbf{0}$, since we already understand in that case how the recursion leads to a rational generating function. The **boundary value locations** are any $B \subseteq (\mathbb{Z}^+)^d$ closed under $\leq$. In the examples below, B will always be the singleton $\{\mathbf{0}\}$. The **boundary values** are a set of values $\{b_{\mathbf{r}} : \mathbf{r} \in B\}$. We study the initial value problem with initial conditions

$$a_{\mathbf{r}} = b_{\mathbf{r}} \text{ for all } \mathbf{r} \in B \tag{2.6}$$

and with the recursion (2.4) assumed to hold for all $\mathbf{r} \in (\mathbb{Z}^+)^d \setminus B$, and with the convention that summands with $\mathbf{r} - \mathbf{s} \notin (Z^+)^d$ are zero. Thus the data for the problem is E, Q, B and $\{b_{\mathbf{r}} : \mathbf{r} \in B\}$. Figure 2.1 shows an example of this with $E = \{(2,-1),(1,-2)\}$, with B taken to be the y-axis.

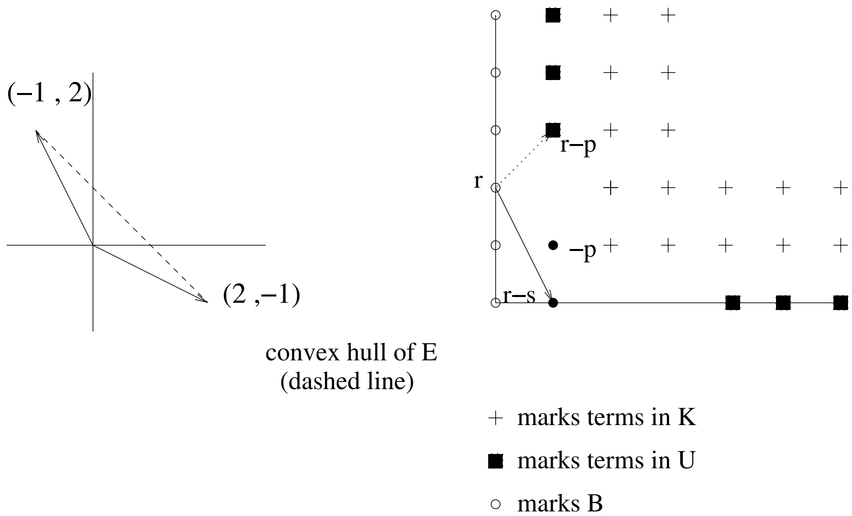

Figure 2.1: The set E and a Newton diagram of K and U

Let Z denote the set $(\mathbb{Z}^+)^d \setminus B$ and 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 (2.4) holds. It is more convenient to work with F_Z and then recover F from F_Z via $F = F_Z + F_B$ where

$$F_B = \sum_{\mathbf{r} \in B} b_\mathbf{r} \mathbf{z}^\mathbf{r} .$$

To apply the kernel method, one examines the product QF_Z. There are two kinds of contribution to QF_Z. Firstly, 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 $c_\mathbf{s} b_{\mathbf{r}-\mathbf{s}} \mathbf{z}^{\mathbf{r}-\mathbf{p}}$ coming from the difference between the coefficient of $\mathbf{z}^{\mathbf{r}-\mathbf{p}}$ in QF_Z and the coefficient in QF, which vanishes. 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. The example in Figure 2.1 has terms of K in the first two rows and columns of Z. Secondly, for every pair $(\mathbf{r}, \mathbf{s})$ with $\mathbf{s} \in E, \mathbf{r} - \mathbf{s} \in Z$ and $\mathbf{r} \notin Z$, there is a term $-c_\mathbf{s} a_{\mathbf{r}-\mathbf{s}} \mathbf{z}^{\mathbf{r}-\mathbf{p}}$ coming from 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. In the example from Figure 2.1, U has one row and one column of terms; the value of $\mathbf{r}$ leading to the xy^3-term of U is pictured.

Lemma 2.15 ([BMP00, Theorem 5]). *Let E be a finite subset of $(\mathbb{Z}^+)^d$ whose convex hull does not intersect the negative orthant. Let $\{c_\mathbf{s} : \mathbf{s} \in E\}$ be constants, let $\mathbf{p}$ be the coordinatewise infimum* *of E as above, let $B \subseteq (\mathbb{Z}^+)^d$ be closed under $\leq$, let $\{b_{\mathbf{s}} : \mathbf{s} \in B\}$ be constants and let*

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

Then there is a unique set of values $\{a_{\mathbf{r}} : \mathbf{r} \in Z\}$ such that (2.4) holds for all $\mathbf{r} \in Z$. Consequently, there is a unique pair of formal power series F_Z and U such that

$$QF_Z = K - U .$$

Furthermore, if K is analytic in a neighborhood of the origin, then so are U and F_Z. □

Proof: The convex hull of E and the closed negative orthant are disjoint convex polyhedra so there is a hyperplane that separates them and meets neither. The normal vector may be perturbed slightly to obtain a rational vector $\mathbf{v}$ such that $\mathbf{v} \cdot \mathbf{s} > 0$ for all $\mathbf{s} \in E$ and $\mathbf{v} \cdot \mathbf{s} < 0$ for all $\mathbf{s} \neq \mathbf{0}$ in the negative orthant. The vector $\mathbf{v}$ must have positive coordinates. Clearing denominators, we may assume $\mathbf{v}$ is an integer. Linearly order $(\mathbb{Z}^+)^d$ by the value of the dot product with $\mathbf{v}$, breaking ties arbitrarily, to produce a well ordering, $\preceq$, of $(\mathbb{Z}^+)^d$ and hence of Z.

Now proceed by induction on $\mathbf{r}$ with respect to $\preceq$. Fix $\mathbf{m} \in Z$. If $\mathbf{s} \in E$ and $\mathbf{r} \prec \mathbf{m}$ then $\mathbf{r} - \mathbf{s} \prec \mathbf{m}$. Consequently, the validity of (2.4) for all $\mathbf{r} \prec \mathbf{m}$ depends only on values $a_{\mathbf{r}}$ with $\mathbf{r} \prec \mathbf{m}$. Assume for induction that there is a unique set of values of $\{a_{\mathbf{r}} : \mathbf{r} \prec \mathbf{m}\}$ such that (2.4) holds for $\mathbf{r} \prec \mathbf{m}$. Imposing (2.4) for $\mathbf{r} = \mathbf{m}$ then uniquely specifies $a_{\mathbf{m}}$, completing the induction.

To show convergence, let $\gamma' = \log \sum_{\mathbf{s}\in E} |c_{\mathbf{s}}|$. By analyticity of K we may choose $\gamma' \geq \gamma$ for which $|b_{\mathbf{r}}| \leq \exp(\gamma\, \mathbf{r} \cdot \mathbf{v})$. With this as the base step, it follows by induction that this holds for $a_{\mathbf{r}}$ in place of $b_{\mathbf{r}}$:

$$\begin{aligned} |a_{\mathbf{r}}| &\leq \left(\sum_{\mathbf{s}\in E} |c_{\mathbf{s}}|\right) \sup_{\mathbf{s}\in E} |a_{\mathbf{r}-\mathbf{s}}| \\ &\leq e^{\gamma} \sup_{\mathbf{m}\cdot\mathbf{v}<\mathbf{r}\cdot\mathbf{v}} |a_{\mathbf{m}}| \\ &\leq e^{\gamma} e^{\gamma\,(\mathbf{r}\cdot\mathbf{v}\,-\,1)} \\ &= e^{\gamma\,\mathbf{r}\cdot\mathbf{v}} , \end{aligned}$$

establishing an exponential bound on $|a_{\mathbf{r}}|$ and hence analyticity of F near the origin. From this, analyticity of F_Z and U follow. □

The previous lemma is based on a formal power series approach. 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 whatever factor of Q vanishes at the origin. The kernel method turns this intuition into a precise statement.

Theorem 2.16 ([BMP00, Theorem 13]). *Let $d \geq 2$ be arbitrary and suppose the boundary locations B are of the form $\mathbf{r} : \mathbf{r} \not\geq \mathbf{s}\}$ for some $\mathbf{s} \in (\mathbb{Z}^+)^d$. If $p_1, \ldots, p_{d-1} \geq 0 > p_d$ and the boundary generating function $K(\mathbf{z})$ is algebraic, then F is algebraic.*

PROOF: Suppose $\mathbf{r} \notin Z$ and $\mathbf{r} - \mathbf{s} \in Z$ with $\mathbf{s} \in E$. We know that $\mathbf{r} - \mathbf{s}' \notin Z$, where $\mathbf{s}'_j = \mathbf{s}_j$ for all $j \leq d-1$ and $\mathbf{s}'_d = 0$; this is because the complement of Z is closed under coordinatewise $\leq$ and the first $d-1$ coordinates of any point in E are nonnegative. Thus $s_d - p_d < r_d \leq s_d$. It follows that U is $x_d^{s_d+1}$ a polynomial of degree at most $p_d - 1$ in x_d.

The polynomial Q is equal to $z_d^{p_d} - \sum_{\mathbf{s} \in E} c_{\mathbf{s}} z_d^{p_d} \mathbf{z}^{\mathbf{s}}$. It is convenient to regard this as a polynomial in z_d over the field of algebraic functions of $z_1, \ldots, z_{d-1}$. The degree of Q in z_d is at least p_d. Let $\{\xi_i(z_1, \ldots, z_{d-1})\}$ be the roots of this polynomial. At least p_d of these, when counted with multiplicities, satisfy $\xi_i(\mathbf{0}) = 0$: this follow from the fact that $(0, \ldots, 0, j) \notin E$ for any negative j, whence the polynomial $Q(0, \ldots, 0, z_d)$ has multiplicity p_d at 0.

If the p_d such roots of Q are distinct, then the equation $QF_Z = K - U$ evaluated at each ξ_i leads to p_d equations

$$U(\xi_i) = K(\xi_i)\,.$$

The Lagrange interpolation formula [PS76, Section V1.9] produces a polynomial P given its values $y_1, \ldots, y_k$ at any k points $x_1, \ldots, x_k$:

$$P(x) = \sum_{j=1}^{n} y_j \prod_{i \neq j} \frac{x - x_i}{x_j - x_i}\,. \tag{2.7}$$

Over any field of characteristic zero, and in particular over the algebraic functions of $x_1, \ldots, x_d$, this is the unique polynomial of degree at most $k-1$ passing through the k points. Taking $k = p_d$, $x_i = \xi_i$, and $y_i = K(\xi_i)$ shows that U is given by (2.7). Thus U a rational function of algebraic functions and is therefore algebraic. Finally, if the ξ_i are not distinct, one has instead the p_d equations:

$$U(\xi_i) = K(\xi_i),\; U'(\xi_i) = K'(\xi_i),\; \ldots, U^{(m_i-1)}(\xi_i) = k^{(m_i-1)}(\xi_i)$$

where m_i is the multiplicity of the root ξ_i. One may replace the Lagrange interpolation formula by the Hermite interpolation formula [IK66, Section 6.1, Problem 10], which again gives U as a rational function of each $K(\xi_i)$ and its derivatives. $\square$

Specializing further to $d = 2$ and $B = \{\mathbf{0}\}$ gives the following explicit formula for F.

Corollary 2.17 ([BMP00, equation (24)]). *Suppose further that $d = 2$, $\mathbf{p} = (0, -p)$, and $B = \{\mathbf{0}\}$ with boundary value $b_{\mathbf{0}} = 1$. 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 $Q(x, y) = -C(x) \prod_{j=1}^{p}(y - \xi_j(x)) \prod_{j=1}^{P}(y - \rho_j(x))$ for some r and $\rho_1, \ldots, \rho_P$. The generating function F_Z will then be given by*

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

$\square$

PROOF: Work in the ring $\mathbb{C}[[x]][y]$ of polynomials in y with coefficients in the local ring of power series in x converging in a neighborhood of zero. The asserted factorization of Q follows from its vanishing to order p at $y = 0$ and having having degree $p + r$ (as a polynomial in y with coefficients in $\mathbb{C}[[x]]$). By definition, $K(x, y) = y^p$. The degree in y of $U(x, y)$ is at most $p-1$, since ... It follows that the degree of $K(x, y) - U(x, y)$ in y is exactly p. If we know p factors $(y - a_j)$ and the leading coefficient C of a polynomial of degree p, then the polynomial is completely determined: it must be $C \prod_{j=1}^{p}(y - a_j)$. Since $K - U$ vanishes on $y = \xi_j(x)$ in a neighborhood of zero for all j, it is divisible by $\prod_{j=1}^{p}(y - \xi_j(x))$. The leading coefficient of $K - U$ is the same as the leading coefficient of K, namely 1. Therefore, $K - U = \prod_{j=1}^{p}(y - \xi_j(x))$, which establishes the conclusion of the corollary. $\Box$

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** from $(0, 0)$ to (r, s) with increments in E are the paths which never go below the horizontal axis.

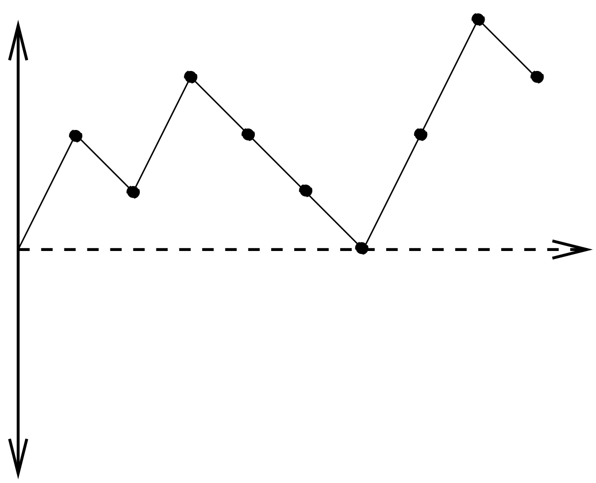

Figure 2.2: a generalized Dyck path of length nine with $E = \{(1, 2), (1, -1)\}$

Let $F(x, y) = \sum_{r,s} a_{rs} x^r y^s$ generate the number, a_{rs}, of generalized Dyck paths to the point (r, s). In the notation of the previous discussion, we have $\mathbf{q} = (0, 0)$, $F = F_q$, $Q(x, y) = y^p(1 - \sum_i x^{r_i} y^{s_i})$, and $C(x) = \sum_{i:s_i = P} x^{r_i}$. The special case $p = P = 1$, that is, vertical displacement of at most 1 per step, occurs often in classical examples.

Proposition 2.18. *Let $E, p, P, C(x)$ be as above and suppose that $p = P = 1$. Then the generating function for generalized Dyck paths with steps from E is given by*

$$F(x, y) = \frac{\xi(x)}{a(x) - C(x)\xi(x)y}$$

where $a(x) = \sum_{i:s_i = -1} x^{r_i}$.

PROOF: Here, Q is quadratic in y and we may simplify the formula for F as follows. The product $\xi\rho$ equals $a(x)/C(x)$ where $a(x) = \sum_{i:s_i=-1} x^{r_i}$, and hence

$$F(x,y) = \frac{\xi(x)/a(x)}{1 - [C(x)\xi(x)/a(x)]y}.$$

□

We now discuss the three standard examples from [BMP00].

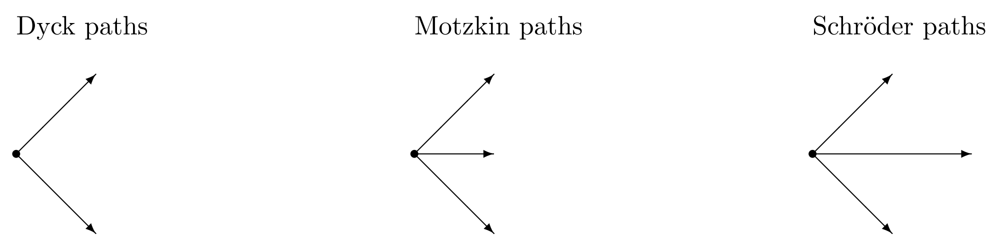

Figure 2.3: legal steps for three types of paths

Dyck paths:

When $E = \{(1,1),(1,-1)\}$ we have the original Dyck paths. We have $p = 1 = P$ and $Q(x,y) = y - xy^2 - x$. 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.

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 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

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

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 $(1 - x^2 \pm \sqrt{1 - 6x^2 + x^4})/(2x)$. Again,

$$F(x,y) = \frac{\xi(x)/x}{1 - y\xi(x)} = \frac{2}{1 + \sqrt{1 - 6x^2 + x^4} - x^2 - 2xy}.$$

2.4 D-finite generating functions

The more explicitly a generating function is described, the better are the prospects for getting information out of it. This includes not only asymptotic estimation, but also proving bijections and, in general, relating the class being counted to other combinatorial classes. Rational generating functions are easy to work because they are specified by finite data: both numerator and denominator are a finite sum of monomials with integer exponents and (usually) integer coefficients.

We have seen that some common and very natural combinatorial operations take us from the class of rational functions to the larger class of algebraic generating functions. These also have canonical representations: if f is algebraic then there is a minimal polynomial P for which $P(f) = 0$; f may be specified by writing down the coefficients of P, which are themselves polynomials and therefore finitely specified. In Section 7.1 we will discuss techniques in computational algebra that allow one to manipulate algebraic functions by performing manipulations directly on the minimal polynomials. This makes the class of algebraic generating functions quite nice to work with. There are, however, common combinatorial operations that take us out of the class of algebraic functions, and this drives us to consider one further step in the hierarchy. A more complete discussion of this hierarchy for

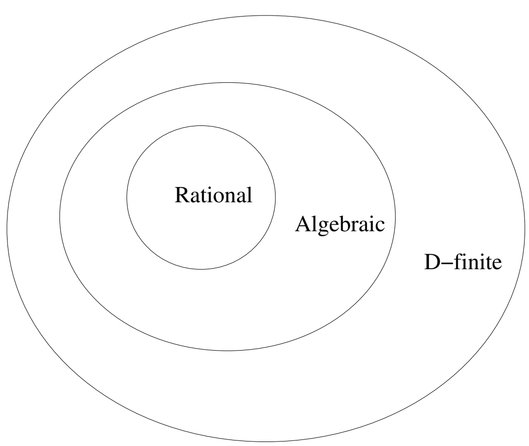

Figure 2.4: some classes of generating functions

univariate functions may be found in [Sta99, Chapter 6]. I will give a brief summary here, first for univariate functions and then for multivariate functions.

Univariate D-finite functions

Some notation, taken from [Sta99], will come in handy. We have already been using $\mathbb{C}[z]$ and $\mathbb{C}[[z]]$ to denote respectively the polynomials and formal power series over $\mathbb{C}$. In order to discuss algebraic and D-finite functions, it is more convenient to work over a field. Denote by $\mathbb{C}(z)$ the field of fractions of $\mathbb{C}[z]$, which are just the rational functions. Denote by $\mathbb{C}((z))$ the field of fractions of $\mathbb{C}[[z]]$; since $\mathbb{C}[[z]]$ has the unique maximal ideal $< z >$, its fraction field is just the formal Laurent series $\mathbb{C}[[z]][1/z]$. The ring of algebraic formal power series is then defined to be the set $\mathbb{C}_{\mathrm{alg}}[[z]]$ of elements of $\mathbb{C}[[z]]$ that are algebraic over $\mathbb{C}(z)$. In other words, clearing denominators, we have $f \in \mathbb{C}_{\mathrm{alg}}[[z]]$ if and only if f is a formal power series and

$$\sum_{j=0}^{m} P_j f^j = 0$$

for some m and some $P_0, \ldots, P_m \in \mathbb{C}[z]$. Equivalently, f is algebraic if and only if the powers $1, f, f^2, \ldots$ span a finite dimensional vector space in $\mathbb{C}((z))$ over $\mathbb{C}(z)$.

Given a formal power series $f = \sum_{n=0}^{\infty} a_n z^n$, the formal derivative f' is defined, as one expects, to be $\sum_{n=0}^{\infty} (n+1) a_{n+1} z^n$. If f is analytic in an open neighborhood $\mathcal{N}$ of zero, then f' is analytic on $\mathcal{N}$ as well. We have not discussed a combinatorial interpretation for differentiation, but it is clear that differentiation will arise from the operation "multiply the n^{th} term by n. Shortly, we will generalize this operation to **polynomial recursion**.

Definition 2.19. *A formal power series $f \in \mathbb{C}[[z]]$ is D-finite if and only if there is an integer m and poloynomials $P, P_0, \ldots, P_m \in \mathbb{C}[z]$ with $P_m \neq 0$ such that*

$$P + P_0 f + P_1 f' + \cdots + P_m f^{(m)} = 0\,. \tag{2.8}$$

Equivalently, f is D-finite if and only if f and its derivatives span a finite dimensional vector space over $\mathbb{C}(z)$.

Variants of this definition are discussed in [Sta99, Proposition 6.4.1]. The Venn diagram depicting the hierarchy in figure 2.4 is justified by the following proposition.

Proposition 2.20 ([Sta99, Theorem 6.4.6]). *If f is algebraic then f is D-finite.*

Proof: By definition, since f is algebraic, there is a polynomial $P = \sum_{j=0}^{m} P_j y^j$ in $\mathbb{C}[z, y]$ for which $P(z, f) = 0$. Implicit differentiation yields

$$f' = -\frac{\partial P(z,y)/\partial z|_{y=f}}{\partial P(z,y)/\partial y|_{y=f}}\,. \tag{2.9}$$

To justify this formally, assume P to be of minimal degree in y; then $\partial P(z,y)/\partial y$ is of lesser degree in y so it is nonzero when evaluated at $y = f$ and taking the derivative of the equation $P(z,f) = 0$ shows that $\partial P(z,y)/\partial z + f' \partial P(z,y)/\partial z$ vanishes on $y = f$, justifying (2.9).

We have shown that f' is in the field of fractions of the ring $\mathbb{C}[z][f]$. The quotient rule now implies that the derivative of any element of $\mathbb{C}(z,f)$ is again in $\mathbb{C}(z,f)$. By induction, all derivatives of f are in $\mathbb{C}(z,f)$. But this is a finite extension of $\mathbb{C}(z)$ for any algebraic f. Thus f and its derivatives span a finite vector space over $\mathbb{C}(z)$, finishing the proof. □

Recall that a series $f = \sum_{n=0}^{\infty} a_n z^n$ is rational if and only if the sequence $\{a_n : n \geq 0\}$ satisfies a linear recurrence with constant coefficients. There is no such quick characterization of coefficients of algebraic generating functions, but there is for D-finite functions, which is another reason the D-finite generating functions are a natural class.

Definition 2.21 (P-recursiveness (1 variable)). *A sequence $\{a_n : n \geq 0\}$ is said to be* ***P-recursive*** *(short for "polynomially recursive") if there exist polynomials $P_0, \ldots, P_m$ with $P_m \neq 0$ such that*

$$P_m(n)a_{n+m} + P_{m-1}(n)a_{n+m-1} + \cdots + P_0(n)a_n = 0 \tag{2.10}$$

for all $n \geq 0$.

Example 2.22. Let $a_n = 1/n!$. Then $(n+1)a_{n+1} - a_n = 0$ for all $n \geq 0$, so $\{a_n : n \geq 0\}$ is P-recursive. The generating function for $(n+1)a_{n+1}$ is f', so we have $f' - f = 0$ showing that f is D-finite. This differential equation may be solved, leading to $f = a_0 e^{-z}$.

The connection between P-recursion and D-finiteness illustrated in the previous example is generalized by the following result, which is stated as [Sta99, Proposition 6.4.3] and attributed to [Com64]. The proof consists of matching up coefficients.

Theorem 2.23 (D-finite ⇔ P-recursive). *A sequence $\{a_n : n \geq 0\}$ is P-recursive if and only if its generating function $f = \sum_{n=0}^{\infty} a_n z^n$ is D-finite.* □

Proof: First suppose f is D-finite. Let $\{P_k : 0 \leq k \leq m\}$ be as in (2.8) and let $b_{k,j}$ denote the x^j-coefficient of P_k. The x^{n-k} coefficient of $f^{(j)}$ is equal to $(n-k+j)_j a_{n-k+j}$ where $(u)_j := u(u-1)\cdots(u-j+1)$ denotes the falling factorial. Equating the coefficient of x^n to zero in the left-hand side of (2.8) gives

$$\sum_{j=0}^{m} \sum_{k} b_{k,j} (n-k+j)_j a_{n-k+j} = 0 \,.$$

This is a linear equation in $\{a_{n+j} : j \in [a,b]\}$ for some finite interval $[a,b]$ whose coefficients are polynomials in n of degrees at most m. It does not collapse to $0 = 0$ because for any j such that $b_{m,j} \neq 0$, the coefficient of a_{n-j+m} is $b_{m,j} n^m + O(n^{m-1}) \neq 0$. We may re-index so that $a = 0$.

Conversely, suppose that (2.10) is satisfied. The polynomials $\{(n+j)_j : j \geq 0\}$ form a basis for $\mathbb{C}[n]$ which is triangular with respect to the basis $\{n^j\}$, hence each P_k is a finite linear combination $\sum c_{k,j}(n+j)_j$. Plugging this into (2.10) yields

$$\sum_{k=0}^{m} \sum_{j=0}^{\deg P_k} c_{k,j}(n+j)_j a_{n+k} = 0 \, .$$

The rules for differentiating formal power series extend to formal Laurent series $C((x))$, in which $(n+j)_j a_{n+k}$ is just the x^n coefficient of $(f \cdot x^{j-k})^{(j)}$. Thus,

$$\sum_{k=0}^{m} \sum_{j=0}^{\deg P_k} c_{k,j}(f \cdot x^{j-k})^{(j)} = 0$$

in $\mathbb{C}((x))$. Using the product rule, this becomes a nontrivial linear ODE in f with coefficients in $\mathbb{C}[x][x^{-1}]$, and multiplying through by a sufficiently high power of x gives a relation of the form (2.8).
□

Multivariate D-finite functions

Let $\mathbb{C}[\mathbf{z}], \mathbb{C}(\mathbf{z})$ and $\mathbb{C}[[\mathbf{z}]]$ denote respectively the polynomials, rational functions and formal power series in d variables, $z_1, \ldots, z_d$. Formal partial differentiation is defined in the obvious way: if $f = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ then $\partial f / \partial z_j := \sum_{\mathbf{r}} (r_j + 1) a_{\mathbf{r}+\delta_j} \mathbf{z}^{\mathbf{r}}$ where δ_j has a 1 in position j and zeros elsewhere. Generalizing the univariate definition, a power series $f \in \mathbb{C}[[\mathbf{z}]]$ is said to be D-finite if f and all its iterated partial derivatives generate a finite dimensional vector space over $\mathbb{C}(\mathbf{z})$.

The correct analogue of P-recursiveness in the multivariate case is not so obvious. The following definition is Definition 3.2 in [Lip89]. Note that the definition is recursive in the dimension, d.

Definition 2.24 (P-recursiveness (d variables)). *Suppose P-recursiveness has been defined for arrays of dimension $d-1$. Then the array $\{a_{\mathbf{r}} : \mathbf{r} \in (\mathbb{Z}^+)^d\}$ is said to be P-recursive if there is some positive integer k such that the following two conditions hold.*

1. *For each $j \in [d]$ there are polynomials $\{P_\nu^j : \nu \in [k]^d\}$, not all vanishing, such that*

$$\sum_{\nu \in [k]^d} P_\nu^j (r_j) a_{\mathbf{r}-\nu} = 0$$

   as long as $\mathbf{r} \geq (k, k, \ldots, k)$ coordinatewise.

2. *All the $(d-1)$-variate arrays obtained from $\{a_{\mathbf{r}}\}$ by holding one of the d indices fixed at a value less than k are P-recursive.*

The following result extends the connection between P-recursiveness and D-finiteness to the d-variate setting. The proof, which I will omit, may be found in [Lip89].

Theorem 2.25. *The array $\{a_{\mathbf{r}} : \mathbf{r} \in (\mathbb{Z}^+)^d\}$ is P-recursive if and only if the generating function $f(\mathbf{z}) := \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is D-finite.* □

Diagonals

D-finite generating functions are finitely specifiable and the arrays they generate satisfy nice recursions. There is one more reason to exapnd our horizons to this class of function, namely its closure properties. The classes of rational and algebraic generating functions are both closed under the ring operations. It will come as no surprise that the D-finite functions are closed under these as well.

Theorem 2.26. *If f and g are D-finite then so are $f + g$ and fg.*

PROOF: Let V be the vector subspace of $\mathbb{C}((x))$ spanned over $\mathbb{C}(x)$ by $f + g$ and all its derivatives. Clearly V is contained in the sum of subspaces $V_f + V_g$ spanned by the derivatives of f and g respectively, hence V is finite dimensional and in fact the dimension is bounded by $\dim(V_f)+\dim(V_g)$.

Let $f^{(\mathbf{r})}$ denote the partial derivative of f taken r_i times with respect to z_i for each $1 \le i \le d$. The products $f^{(\mathbf{r})}g^{(\mathbf{s})}$ span a finite dimensional space, in fact a space of dimension at most $\dim(V_f) \cdot \dim(V_g)$. By the product rule, every derivative $(fg)^{\mathbf{r}}$ is in this space. □

A more interesting and nontrivial closure property has to do with diagonals. The **diagonal** of the bivariate power series $F(x, y) := \sum_{r,s=0}^{\infty} a_{rs}x^r y^s$ is the univariate series $h(z) := \sum_{n=0}^{\infty} a_{n,n}z^n$. [Don't be confused: $h(z)$ is not equal to $F(z, z)$.] This may be generalized to any number of variables. A (generalized) diagonal of a d-variate series $F(\mathbf{z}) := \sum_{\mathbf{r}} a_{\mathbf{r}}\mathbf{z}^{\mathbf{r}}$ is the t-variate series $F_\pi(\mathbf{z}) := \sum_{\mathbf{s}} a_{\pi^{-1}\mathbf{s}}\mathbf{z}^{\mathbf{s}}$ where $\pi : [d] \to [t]$ is any surjection. When $t = 1$ the resulting univariate series is called the **complete** diagonal; when $d = 2$ and $t = 1$ this is the bivariate diagonal previously defined.

A result appearing in [HK71] and credited to [Fur67] is that the diagonal of a bivariate rational power series is always algebraic. This result, though it does not solve the general bivariate asymptotic problem, is handy when one is only interested in the main diagonal. This occurs more often that you might think. For example, the Lagrange inversion formula represents the implicitly defined function f solving $f(z) = z\phi(f(z))$ as the diagonal of the series $y/(1 - x\phi(y))$, with the z_n coefficient then divided by n.

Because the Hautus-Klarner-Fursteberg diagonal extraction method is constructive, it is computationally useful and I will present a proof and an example shortly. To finish the story, though, it turns out that this result is somewhat limited. For one thing, it cannot be used iteratively, since the result does not apply to algebraic generating functions. Secondly, a rational function in more than two variables need not have algebraic diagonals. For example, the generating function for the trinomial coefficients $\binom{3n}{n\,n\,n}$ is the complete diagonal of $1/(1-x-y-z)$, and is not algebraic (see [Sta99, Problem 6.3]). In 1988, Lipshitz proved that D-finite series are closed under taking diagonals:

Theorem 2.27 (Lipshitz 1988). *Any diagonal of a D-finte series is D-finite.*

This is the main result of [Lip88] and the proof of this, though not long, is a little too much to reproduce here. This is, in some sense, the final word on the hierarchy in figure 2.4. A consequence is that the **Hadamard product** of two d-variate D-finite power series is again D-finite. The Hadamard product of $\sum_{\mathbf{r}} a_{\mathbf{r}}\mathbf{z}^{\mathbf{r}}$ with $\sum_{\mathbf{r}} b_{\mathbf{r}}\mathbf{z}^{\mathbf{r}}$ is simply the function $\sum_{\mathbf{r}} a_{\mathbf{r}} b_{\mathbf{r}}\mathbf{z}^{\mathbf{r}}$. The Hadamard product of F and G is a generalized diagonal of $F(\mathbf{x})G(\mathbf{y})$, whence closure under Hadamard products follows from closure under generalized diagonals. Lest the class of D-finite functions seem too good to be true, I should point out that it is not closed under composition. It is shown in [Sta99] that $F \circ G$ is D-finite if F is D-finite and G is algebraic, but not the other way around.

I owe you a proof that the diagonal of a bivariate rational series is algebraic. Stanley [Sta99] gives a formal power series proof but it relies on some results we have not developed on the algebraic closure of $\mathbb{C}[[x]]$ (Puiseux's Theorem). The following analytic proof has the advantage of providing computational information.

Theorem 2.28 (Hautus-Klarner-Furstenberg diagonal extraction). *Let*

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

where P and Q polynomials. Then $h(z) := \sum_{n\geq 0} a_{nn} z^n$ is algebraic.

PROOF: A rational power series converges in a neighborhood of the origin. Hence, when $|y|$ is sufficiently small, the function $F(z, y/z)$ is absolutely convergent for z in some annulus $A(y)$. Treating y as a constant, we view $F(z, y/z)$ as a Laurent series in z inside the annulus $A(y)$; the constant term C of this series is equal to $h(y)$. Thus if we are able to evaluate this constant term as a function of y, we will have the function whose power series in a neighborhood of 0 is h.

By Cauchy's integral formula, $h(y)$ is equal to

$$\frac{1}{2\pi i} \int_{\mathbb{C}} \frac{P(z, y/z)}{zQ(z, y/z)}\, dz$$

where $\mathbb{C}$ is any circle in the annulus of convergence $A(y)$. By the Residue Theorem,

$$h(y) = \sum \text{Res}\left(\frac{P(z,y/z)}{zQ(z,y/z)}; \alpha\right)$$

where the sum is over residues at poles α inside the inner circle of the annulus $A(y)$. The residues are all algebraic functions of y, so we have represented h as the sum of algebraic functions of y. $\square$.

Remark. When computing, one needs to know which poles are inside the circle. They are precisely those that converge to zero as $y \to 0$.

Example 2.29 (Delannoy numbers continued). Recall that $F(z, w) = 1/(1 - z - w - zw)$, so

$$z^{-1} F(z, y/z) = \frac{1}{z - z^2 - y - yz}\,.$$

The poles of this are at

$$z = \frac{1-y}{2} \pm \frac{1}{2}\sqrt{1 - 6y + y^2}\,.$$

Let α_1 denote the root going to zero with y, that is, the one with the minus sign, and let α_2 denote the other root. Since $z^{-1}F(z, y/z) = -1/[(z-\alpha_1)(z-\alpha_2)]$, the residue at $z = \alpha_1$ is just $-1/(\alpha_1 - \alpha_2)$ which is simply $(1 - 6y + y^2)^{-1/2}$. Thus

$$h(y) = \frac{1}{\sqrt{1 - 6y + y^2}} . \tag{2.11}$$

We close the section on D-finite functions with some good news and some bad news concerning diagonal extraction. The good news is that some recent work of Chyzak and Salvy [CS98] has made the extraction of diagonals effective. That is, one may feed a rational function, F, into a black box, along with a specification of which diagonal to compute, and out will pop the polynomial coefficients of a differential equation witnessing the D-finiteness of the diagonal. One may then obtain asymptotics, and indeed for diagonal directions, this is probably the best way to obtain asymptotics. The bad news is that, while the method may be adapted to other diagonals with rational slopes by means of the substitution $F(x^p, y^q)$, the complexity of the computation increases with p and q so there is no way to take limits, no uniformity, and hence no way to use this method to obtain asymptotics that are truly bivariate (allowing both indices to vary simultaneously).

2.5 Exponentiation: set partitions

Let exp denote the power series $\sum_{n=0}^{\infty} z^n$. The exponential e^F of a formal power series in any number of variables may be defined as $\exp \circ F$; this is well defined as long as $F(\mathbf{0}) = 0$, with formal composition defined by (2.1). Exponentiation turns out to have a very useful combinatorial interpretation.

Let $\mathcal{B}$ be a combinatorial class with $b_n := |\mathcal{B}_n|$. Define a **$\mathcal{B}$-partition** of $[n]$ to be a set of pairs $\{(S_\alpha, G_\alpha) : \alpha \in I\}$, where the collection $\{S_\alpha : \alpha \in I\}$ is a partition of the set $[n]$ and each G_α is an element of $\mathcal{B}_{|S_\alpha|}$. Define the class $\exp(\mathcal{B})$ to be the class of $\mathcal{B}$-partitions enumerated by n, that is, $\exp(\mathcal{B})_n$ is the class of $\mathcal{B}$-partitions of $[n]$.

Example 2.30. Take $\mathcal{B}$ to be the class of connected graphs with labelled vertices, enumerated by number of vertices. Given $S \subseteq [n]$ and a graph G with $|S|$ vertices, labelled $1, \ldots, |S|$, let $\langle S, G\rangle$ denote the graph G with each label j replaced by s_j, where $S = \{s_1 < \cdots < s_{|S|}\}$. Replacing each pair (S_α, G_α) by $\langle S, G\rangle$, we have an interpretation of $\exp(\mathcal{B})_n$ as a collection of connected graphs whose labels are $1, \ldots, n$, each used exactly once. In other words, the exponential of the class of labelled connected graphs is the class of all labelled graphs.

The use of the word "exponential" for the combinatorial operation described above is justified by the following theorem.

Theorem 2.31 (exponential formula). *Let $g(z)$ be the exponential generating function for the class $\mathcal{B}$, that is, $g(z) = \sum_{n=0}^{\infty} \frac{b_n}{n!} z^n$. Then $\exp(g(z))$ is the exponential generating function for $\exp(\mathcal{B})$*

and $\exp(y\, g(z))$ *is the* ***semi-exponential generating function*** *whose* $y^k z^n$*-coefficient is* $1/n!$ *times the number of elements of* $\exp(\mathcal{B})_n$ *with* $|I| = k$*.*

PROOF: Let $\mathcal{A}$ denote the bivariate class $\exp(\mathcal{B})$ broken down by $|I|$, so that $a_{nk} := |\mathcal{A}_{nk}|$ counts elements of $\exp(\mathcal{B})_n$ with $|I| = k$. Since every set of size k may be listed in $k!$ different orders, we see that $a_{nk} = u_{nk}/k!$, where u_{nk} counts sequences of pairs, $((S_j, G_j), \ldots, (S_k, G_k))$ of partitions of $[n]$ and associated elements of $\mathcal{B}$.

Just as $g(z)$ is the exponential generating function for the class $\mathcal{B}$, it is also the ordinary generating function for the class $\mathcal{B}$ re-weighted so that each element of $\mathcal{B}_k$ counts with weight $1/k!$. Therefore, $g(z)^k$ counts sequences of length k with the weight of a sequence given by $w(G_1, \ldots, G_k) := \prod_{j=1}^k (1/|G_j|!)$. The reason it is useful to count these sequences by total weight is that every such sequence appears exactly $\binom{n}{|G_1| \cdots |G_k|}$ times as the sequence of second coordinates of sequences of pairs counted by u_{nk}. This multinomial coefficient is exactly $n!\, w(G_1, \ldots, G_k)$, so we see that

$$u_{nk} = n![y^k z^n](y\, g(z))^k .$$

The relation between u_{nk} and a_{nk} yields

$$\frac{a_{nk}}{n!} = [y^k z^n] \frac{(y\, g(z))^k}{k!} .$$

Since $(y\, g(z))^k$ has a y^j-coefficient only when $j = k$, we may sum the right-hand side:

$$\frac{a_{nk}}{n!} = [y^k z^n] \sum_{k=0}^{\infty} \frac{(y\, g(z))^k}{k!} = \exp(y\, g(z)) .$$

Thus $\exp(y\, g(z))$ is the semi-exponential generating function for $\{a_{nk}\}$ (the ordinary generating function for $a_{nk}/n!$). This proves the second claim. The first follows from setting $y = 1$. □

To use the exponential formula, one needs exponential generating functions to input.

Example 2.32 (egf for permutations). The number of permutations of $[n]$ is $n!$, so the exponential generating function for permutations is

$$f(z) = \sum_{n=0}^{\infty} \frac{n!}{n!} z^n = \frac{1}{1-z} . \tag{2.12}$$

Subtracting 1, the generating function for non-empty permutations is $z/(1-z)$.

Example 2.33 (egf for cycles). A proportion of $1/n$ of permutations of size n are a single n-cycle. This follows, for instance, by computing recursively the probability that $\pi^k(1) = 1$ given $\pi(j)(1) \neq 1 \,\forall\, j \leq k-1$. The exponential generation function for non-empty n-cycles is therefore

$$g(z) = \sum_{n=1}^{\infty} \frac{1}{n} z^n = \log\left(\frac{1}{1-z}\right) . \tag{2.13}$$

Applying the exponential formula to each of these two examples gives:

Example 2.34 (permutations by number of cycles). A permutation is the commuting product of its cycles. Thus the class $\mathcal{A}$ of permutations is the exponential of the class $\mathcal{B}$ of non-empty cycles. By the exponential formula, the relation $f = \exp(g(z))$ holds between them, which agrees with (2.12) and (2.13): $1/(1-z) = \exp(\log(1/(1-z)))$. Enumerating permutations by cycle, we get the exponential generating function

$$f(z) = \exp\left(y \log \frac{1}{1-z}\right) .$$

We may write this compactly as $\frac{1}{(1-z)^y}$, though this is not contentful since we have no definition of the y-power of a series other than exponentiation of y times the logarithm.

The number of permutations of $[n]$ with k cycles is called a **Stirling number of the first kind** and is denoted variously as $\left[\begin{matrix} n \\ k \end{matrix}\right]$, $(-1)^k s(n,k)$, $c(n,k)$ as well as other notations.

Example 2.35 (set partitions). The Stirling number of the second kind, denoted $\left\{\begin{matrix} n \\ k \end{matrix}\right\}$, count partitions of $[n]$ into k nonempty sets. To count partitions $\{S_\alpha : \alpha \in I\}$, set $|\mathcal{B}_n| = 1$ for each $n \geq 1$. The generation function for $\mathcal{B}$ is $\exp(z) - 1$, so the generating function for partitions by number of sets is

$$\exp(y(e^z - 1)) .$$

In particular, the exponential generating function for all partitions is

$$e^{e^z - 1} .$$

Example 2.36 (partitions into ordered sets). Just as permutations are the exponential of the class of cycles, the exponential of the class of non-empty permutations is the class of partitions into ordered sets, that is, collections of sequences $\{(x_{11}, \ldots, x_{1n_1}), \ldots, (x_{k1}, \ldots, x_{kn_k})\}$ where each element of $[n]$ appears exactly once as some $x_{i,j}$. Thus the semi-exponential generating function $F(y,z)$ for partitions into ordered sets by total size number and number of sets is given by exponentiating again

$$F(y,z) = \exp\left(y \frac{z}{1-z}\right) . \tag{2.14}$$

Example 2.37 (involutions). An involution is a permutation whose square is the identity. Equivalently, all its cycles must be of length 1 or 2. Thus the class $\mathcal{A}$ of involutions is the exponential of the class $\mathcal{B}$ of cycles of length 1 or 2, enumerated by length. There is just one of each length, so the exponential generating function g of the class $\mathcal{B}$ class is $z + \frac{z^2}{2}$. Hence the exponential generating function f for the class of involutions is $\exp(z + z^2/2)$.

Example 2.38 (2-regular graphs). A 2-regular graph is a graph with no self-edges in which every vertex has degree 2. A labelled 2-regular graph is the union of labelled, undirected cycles, whence the class of labelled 2-regular graphs is the exponential of the class of labelled undirected cycles. Let $\mathcal{A}$ denote this class. We do not allow parallel edges, so the cycles must have length at least 3. What is the number $a_n := |\mathcal{A}_n|$ of labelled 2-regular graphs on n vertices?

Every undirected cycle of length $n \geq 3$ corresponds to two directed cycles. Counting a permutation π as having weight $w = 2^{-N(\pi)}$ where $N(\pi)$ is the number of cycles, and letting p be the proportion of permutations having no short cycles (cycles of length less than 3), we see that $a_n = n!\, p\overline{w}$, where $\overline{w}$ is the average of w over permutations having no short cycles. It is known that $p = \Theta(1)$ and $N(\pi) \sim \log n$ for all but a vanishing proportion of permutations so it would seem likely that $a_n/n! = \Theta(2^{-\log n}) = \Theta(n^{-\log 2})$. This gives a rigorous lower bound: by convexity of 2^{-x}, the average of 2^{-N} over permutations with no short cycles is at least $2^{-\overline{N}}$. It takes a generating function, however, to correct this to a sharp estimate.

Let $u(z)$ is the exponential generating function for undirected cycles of length at least 3. By (2.13)

$$u(z) = \frac{1}{2}\left(\log \frac{1}{1-z} - z - \frac{z^2}{2}\right) .$$

Applying the exponential formula shows that the exponential generating function for labelled 2-regular graphs is

$$\frac{e^{-\frac{1}{2}z - \frac{1}{4}z^2}}{\sqrt{1-z}} .$$

Methods in the next chapter convert this quickly into a good estimate.

Notes

The transfer matrix method is discussed in [Sta97, Section 4.7] and throughout parts of [GJ83, Chapter 2]. The discussion of the kernel method borrows liberally from [BMP00]. the method itself, which appears to have been re-discovered several times, has been taken much further; see, e.g., [FM77, FH84, FIM99] for some applications involving nontrivial amounts of number theory. My discussion of the exponential formula is inspired by [Wil94]. There, the origin of the exponential formula is attributed to the doctoral work of Riddell [RU53], becoming greatly expanded around 1970 by [BG71] and [FS70].

Most of the proofs in Section 2.4 are taken from [Sta99]. An earlier definition of P-recursiveness appeared in the literature but it was discarded, due to its failure to be equivalent to D-finiteness; counterexamples are given in [Lip89]. Lipshitz's Theorem replaced two earlier proofs with gaps, found in [Ges81] and [Zei82]. It also solved a problem of Stanley [Sta80, question 4e]. One might consider a still larger class of generating function, namely the **differentially algebraic** functions, defined to be those that satisfy an equation

$$P(z, f, f', \ldots, f^{(m)}) = 0$$

for some $m > 0$ and some polynomial P. The question of possible behaviors of the coefficient sequence of such a function is wide open; some of the few known results are in [Rub83, Rub92].

With regard to effective computing within each of these classes, a great deal is known about algebraic computations. Some of this is discussed in Section 7.1 below. For D-finite functions, there has been significant recent progress, a very brief discussion of which is given in Section 7.3. Two good references are [SST00] and [CS98]; see also [CMS05] for computing with symmetric D-finite functions. Little or nothing is known as to effective computability in the class of differentially algebraic functions.

Exercises

Exercise 2.1 (counting domino tilings). Let a_{nk} be the number of ways of placing k non-overlapping dominoes on a $2 \times n$ grid. Find the generating function for these numbers.

Exercise 2.2 (counting almost binary trees). Define a class of "binary until the end" trees, by altering the definition in Example 2.13 so that each vertex must have at most d children *unless* all the children are leaves, in which case an arbitrary number is permitted. Adapt the argument from Example 2.13 to compute a generating function G that counts these trees by the number of vertices.

Exercise 2.3 (P-recursiveness).

1. Find a linear differential equation satisfied by $f(z) := \sum_{n \geq 0} n! z^n$ over $\mathbb{C}[z]$.
2. For which real α is the sequence $a_n := (n!)^\alpha$ P-recursive?
3. Determine whether $f(x, y) := \sum_{r,s \geq 0} (r+s)!^3 x^r y^s$ is D-finite. Is it easier to do this directly or to use the following hint? [Hint: show that $a_{r,s} = (r+s)!^\alpha$ satisfies part (i) of the definition of P-recursion with $k = 1$ for any α. Use this and the previous part of this exercise to determine for which α the array $\{(r+s)!^\alpha\}$ is P-recursive.]

Exercise 2.4 (diagonal of slope 2). Obtain the generating function for the next-simplest diagonal, $a_{i,2i}$, of the Delannoy numbers and compare the algebraic complexity to the main diagonal generating function (2.11).

Exercise 2.5 (D-finite generating function). Let $p_0 = 1$ and define $\{p_N : N \geq 1\}$ recursively by

$$p_N = \frac{1}{3N+1} \sum_{j=2}^{N} p_{j-2} p_{N-j}$$

(the sum is empty when $N = 1$). This is generating function from [LP04] gives the probability that a genome in a certain model due to Kaufmann and Levin cannot be improved by changing one allele. Find a differential equation satisfied by the generating function $f(z) := \sum_{N=0}^{\infty} p_N z^N$. Then use Maple (or its equivalent) to solve this Riccati equation explicitly in terms of Bessel functions. Among the solutions, find the only one that is analytic in a neighborhood of the origin.