Combinatorics

C. Krattenthaler

1 Introduction

Combinatorics is a vast field which enters particularly in a crucial way in Statistical Physics. There, it is particularly enumerative problems which are of importance. Therefore, in this article, we shall mainly concentrate on the enumerative aspects of Combinatorics. Section 2 will recall the basic terminology, in particular the basic combinatorial objects and numbers, together with the simplest facts about them. The next six sections will then provide introductions into the most important techniques of enumeration: the generating function technique (see Section 3), Redfield–Pólya theory (see Section 4), methods of solving functional equations of combinatorial origin (see Section 5), methods of asymptotic enumeration (see Section 6), the theory of heaps (see Section 7), and the transfer matrix method (see Section 8). The subsequent sections then discuss specific problem circles with relation to Statistical Physics more closely. Section 9 discusses lattice path problems, Section 10 explains Kasteleyn's method of enumerating perfect matchings and tilings, Section 11 presents the fundamental theorems on non-intersecting paths, and Sections 12 and 13 provide an introduction into the research field involving vicious walkers, plane partitions, rhombus tilings, alternating sign matrices, six vertex configurations, and fully packed loop configurations. Finally, in Section 14, it is explained how one should treat binomial and hypergeometric series, which frequently arise in enumeration problems.

2 Basic combinatorial terminology

In this section we review basic combinatorial notions and facts. The reader finds a more detailed treatment and further results, for example, in [39, Ch. 1].

The basic combinatorial choice problems, and their solutions, are: There are 2^n subsets of an n-element set. There are $\binom{n}{k}$ k-element subsets of an n-element set. Given an alphabet $\mathcal{A} = \{a_1, a_2, \dots\}$, a *word* is a (finite or infinite) sequence of elements of $\mathcal{A}$. Usually, a finite word is written in the form $w_1 w_2 \dots w_n$ (with $w_i \in \mathcal{A}$). Out of the letters $\{1, 2, \dots, k\}$, one can build k^n words of length n. Out of the letters $\{1, 2, \dots, k\}$, one can build $\binom{n+k-1}{n}$ increasing sequences of length n. The number of permutations of an n-element set is $n!$. The set of permutations of $\{1, 2, \dots, n\}$ is denoted by $\mathfrak{S}_n$. The number of permutations of an n-element set with exactly k cycles is the *Stirling number of the first kind* $s(n, k)$. These numbers are given as the expansion coefficients of falling factorials,

$$x(x-1)\cdots(x-n+1) = \sum_{k=0}^{n} (-1)^{n-k} s(n,k)\, x^k,$$

or in form of the double (formal) power series

$$\sum_{n,k\geq 0} s(n,k) x^k \frac{y^n}{n!} = (1+y)^x.$$

A *partition of a set* is a collection of pairwise disjoint subsets the union of which is the complete set. The subsets in the collection are called the *blocks* of the partition. The total number of partitions of an n-element set is the *Bell number* B_n. These numbers are given by

$$\sum_{n\geq 0} B_n \frac{x^n}{n!} = e^{e^x - 1}.$$

The number of partitions of an n-element set into exactly k blocks is the *Stirling number of the second kind* $S(n, k)$. These numbers are given by

$$\sum_{n,k\geq 0} S(n,k) x^k \frac{y^n}{n!} = e^{x(e^y - 1)},$$

or, explicitly, by

$$S(n,k) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n.$$

A *composition* of a positive integer n is a representation of n as a sum $n = s_1 + s_2 + \cdots + s_k$ of other positive integers s_i, where the order of the summands matters. The total number of compositions of n is 2^{n-1}. The number of compositions of n with exactly k summands is $\binom{n-1}{k-1}$. A *partition* of a positive integer n is a representation of n as a sum $n = \lambda_1 + \lambda_2 + \cdots + \lambda_k$ of other positive integers λ_i, where the order of the summands *does* matter. Thus, we may assume that the summands are ordered, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. This is the motivation to write partitions most often in the form of tuples $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ the entries of which are weakly decreasing. The summands of a partition are called the *parts* of the partition. Let $p(n)$ denote the number of partitions of n. These numbers are given by

$$\sum_{n=0}^{\infty} p(n)\, x^n = \frac{1}{\prod_{i=1}^{\infty}(1-x^i)}.$$

If $p(n,k)$ denotes the number of partitions of n into at most k parts, then we have

$$\sum_{n=0}^{\infty} p(n,k) x^n = \frac{1}{\prod_{i=1}^{k}(1-x^i)}.$$

Finally, if $p(n,k,m)$ denotes the number of partitions of n into at most k parts, all of which are at most m, then

$$\sum_{n\geq 0} p(n,k,m)\, x^n = \frac{(1-x^{k+m})(1-x^{k+m-1})\cdots(1-x^{m+1})}{(1-x^k)(1-x^{k-1})\cdots(1-x)}.$$

The expression on the right-hand side is called *q-binomial coefficient*, and is denoted by $\left[{k+m \atop k}\right]_x$.

Partitions are frequently encoded in terms of their *Ferrers diagrams*. The *Ferrers diagram* of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ is an array of cells with ℓ left-justified rows and λ_i cells in row i. For example, the diagram in Figure 1 is the Ferrers diagram of the partition $(3,3,2)$.

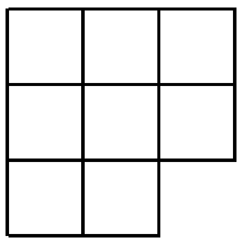


Figure 1: A Ferrers diagram

A *lattice path* P in $\mathbb{Z}^d$ (where $\mathbb{Z}$ denotes the set of integers) is a path in the d-dimensional integer lattice $\mathbb{Z}^d$ which uses only points of the lattice, that is, it is a sequence $(P_0, P_1, \ldots, P_l)$, where $P_i \in \mathbb{Z}^d$ for all i. The vectors $\overrightarrow{P_0P_1}, \overrightarrow{P_1P_2}, \ldots, \overrightarrow{P_{l-1}P_l}$ are called the *steps* of P. The number of steps, l, is called the *length* of P. Figure 2 shows a lattice path in $\mathbb{Z}^2$ of length 11.

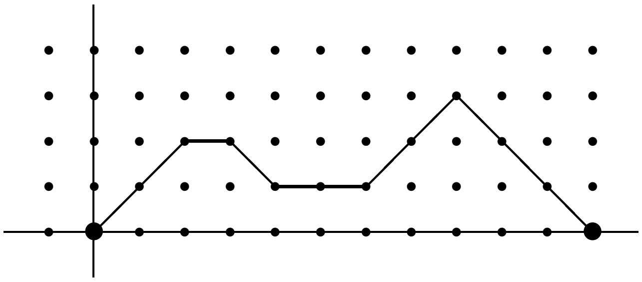


Figure 2: A Motzkin path

A *Dyck path* is a lattice path in the integer plane $\mathbb{Z}^2$ consisting of up-steps $(1, 1)$ and down-steps $(1, -1)$, which starts at the origin, never passes below the x-axis, and ends on the x-axis. See Figure 3 for an example.

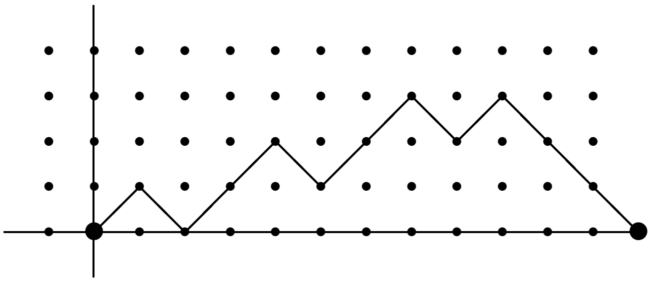


Figure 3: A Dyck path

The number of Dyck paths of length $2n$ is the *Catalan number* $C_n = \frac{1}{n+1}\binom{2n}{n}$. The generating function (see Section 3 for an introduction into the theory of generating functions) for these numbers is

$$\sum_{n=0}^{\infty} C_n\, x^n = \frac{1 - \sqrt{1-4x}}{2x}. \tag{2.1}$$

The reader is referred to Exercise 6.19 in [40] for countless occurrences of the Catalan numbers.

A *Motzkin path* is a lattice path in the integer plane $\mathbb{Z}^2$ consisting of up-steps $(1,1)$, level steps $(1,0)$, and down-steps $(1,-1)$, which starts at the origin, never passes below the x-axis, and ends on the x-axis. The path in Figure 2 is in fact a Motzkin path. The number of Motzkin paths of length n is the *Motzkin number* $M_n = \sum_{k\geq 0} \frac{1}{k+1}\binom{2k}{k}\binom{n}{2k}$. The generating function for these numbers is

$$\sum_{n=0}^{\infty} M_n\, x^n = \frac{1-x-\sqrt{1-2x-3x^2}}{2x^2}. \tag{2.2}$$

The reader is referred to Exercise 6.38 in [40] for numerous occurrences of the Motzkin numbers.

A *Schröder path* is a lattice path in the integer plane $\mathbb{Z}^2$ consisting of horizontal steps $(1,0)$ and vertical steps $(0,1)$, which starts at the origin, never passes below the diagonal $x=y$, and ends on the diagonal $x=y$. See Figure 4 for an example.

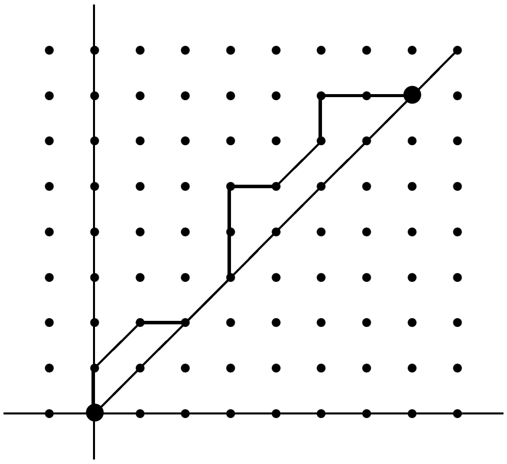


Figure 4: A Schröder path

The number of Schröder paths of length n is the (large) *Schröder number* $S_n = \sum_{k\geq 0} \frac{1}{k+1}\binom{2k}{k}\binom{n+k}{2k}$. The generating function for these numbers is

$$\sum_{n=0}^{\infty} S_n\, x^n = \frac{1-x-\sqrt{1-6x+x^2}}{2x}. \tag{2.3}$$

The reader is referred to Exercise 6.39 in [40] for numerous occurrences of the Schröder numbers.

There is another famous sequence of numbers which we did not touch yet, the *Fibonacci numbers* F_n. They are given by

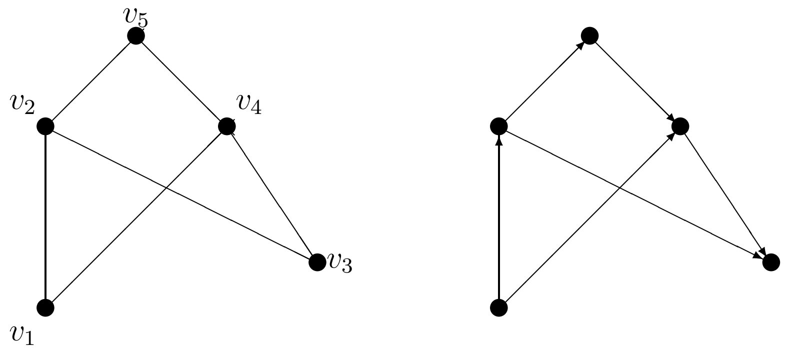
$$F_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n+1},$$

with generating function

$$\sum_{n=0}^{\infty} F_n\, x^n = \frac{1}{1-x-x^2}. \tag{2.4}$$

They also occur in numerous places. For example, the number F_n counts all paths on the integers $\mathbb{Z}$ from 0 to n with steps $(1,0)$ and $(2,0)$.

An *undirected graph* G consists of *vertices* and *edges*. An edge is a 2-element subset of the vertices, which, however, is thought of a line or curve connecting the two vertices. See Figure 5.a for an example. The usual notation for a graph G is $G = (V, E)$, where V is the set of vertices and where E is the set of edges of G. A graph is *planar* if it is embedded in the plane (sphere) in such a way that the curves which mark the edges do not intersect in their interiors. There can be several different ways to embed the same graph in the plane (or in another surface). When we speak of a *planar graph* then we assume the graph already be embedded in a given way. For example, the graph in Figure 5 is not a planar graph, by its drawing. However, there is a different embedding which *is* planar (namely, all embeddings which put the vertex v_3 above the vertex v_5 and leave the other vertices as they are). A *tree* is a graph without any cycles.



a. An undirected graph   b. A directed graph

Figure 5:

A *directed graph* (or *digraph*) G consists of *vertices* and *arcs* (which are sometimes also called *directed edges*). An arc is a pair of vertices, which, however, is thought of an arrow pointing from the first vertex of the pair to the second. See Figure 5.b for an example. The usual notation for a directed graph G is again $G = (V, E)$, where V is the set of vertices and where E is the set of arcs of G. All other notions explained for undirected graphs have analogous meanings for directed graphs.

Graphs can be *labelled*, in which case each vertex is assigned a label, or *unlabelled*. The (undirected) graph in Figure 5.a is labelled, whereas the (directed) graph in Figure 5.b is unlabelled.

3 Generating functions

Generating functions are the very basic tools of enumeration. For introductions into this technique, from different points of view, the reader is referred to [4], [12, Ch. 1], [39, Ch. 1], and [40, Ch. 4].

Let $\mathcal{A}$ be a set of (unlabelled) objects. Each object a in $\mathcal{A}$ has a certain size, $|a|$, which is a non-negative integer. Let us also assume that there is only a finite number of objects from $\mathcal{A}$ of a given size. Let a_n be the number of objects from $\mathcal{A}$ of size n. The (ordinary) *generating function* for $\mathcal{A}$ is the *formal power series*

$$F_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} x^{|a|} = \sum_{n=0}^{\infty} a_n x^n.$$

("Formal" means that x is just an indeterminate, not a real or complex number. One can compute with formal power series in the same way as with analytic series, only that convergence issues do not arise, respectively that "convergence" has a different meaning; cf. [39, Sec. 1.1].) Typical examples are `Sets` (the collection containing all "unlabelled sets", that is all objects of the form $\{\bullet, \bullet, \dots, \bullet\}$, including the empty set, where the size of $\{\bullet, \bullet, \dots, \bullet\}$ is the number of $\bullet$'s) `Sequences` (the collection containing all "unlabelled sequences", that is all objects of the form $(\bullet, \bullet, \dots, \bullet)$, including the empty sequence), `Cycles` (unlabelled cycles), with respective generating function

$$F_{\texttt{Sets}}(x) = F_{\texttt{Sequences}}(x) = \frac{1}{1-x}, \quad F_{\texttt{Cycles}}(x) = \frac{x}{1-x}, \tag{3.1}$$

or `Trees` (unlabelled trees).

If $\mathcal{A}$ and $\mathcal{B}$ are two sets of objects, one can define several other sets of objects using them. The *union* of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A} \cup \mathcal{B}$, has as a groundset the *disjoint* union of $\mathcal{A}$ and $\mathcal{B}$, and the size of an element from $\mathcal{A}$ is its size in $\mathcal{A}$, while the size of an element from $\mathcal{B}$ is its size in $\mathcal{B}$. We have

$$F_{\mathcal{A}\cup\mathcal{B}}(x) = F_{\mathcal{A}}(x) + F_{\mathcal{B}}(x). \tag{3.2}$$

The *product* of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A} \times \mathcal{B}$, has as a groundset the set of pairs $\mathcal{A} \times \mathcal{B}$, and the size of an element (a, b) from $\mathcal{A} \times \mathcal{B}$ is the sum of the sizes of a (in $\mathcal{A}$) and of b (in $\mathcal{B}$). We have

$$F_{\mathcal{A}\times\mathcal{B}}(x) = F_{\mathcal{A}}(x) \cdot F_{\mathcal{B}}(x). \tag{3.3}$$

The substitution of two sets $\mathcal{A}$ and $\mathcal{B}$ of objects can only be defined in certain circumstances, and only in certain more restrictive circumstances the generating function for the substitution can be computed by substituting the generating functions for $\mathcal{A}$ and $\mathcal{B}$. Let us assume that any object a from $\mathcal{A}$ of size n, by its structure, has n atoms (nodes). For example, if $\mathcal{A}$ is a certain set of trees, where the size of a tree is the number of leaves in the tree, then we may take as the atoms the leaves of the tree. In this situation, the *substitution of* $\mathcal{B}$ *in* $\mathcal{A}$, denoted by $\mathcal{A}(\mathcal{B})$, is the set of objects which arises by replacing the atoms of objects from $\mathcal{A}$ by objects from $\mathcal{B}$ in all possible ways. The size of an object from $\mathcal{A}(\mathcal{B})$ is the sum of the sizes of the objects from $\mathcal{B}$ that it contains. In order that $\mathcal{A}(\mathcal{B})$ contains only a finite number of objects of a given size, we must assume that $\mathcal{B}$ contains no elements of size 0. If, in addition, the atoms of any element a from $\mathcal{A}$ inherit an order (for example, if $\mathcal{A}$ is a set of binary trees, then the leaves of a binary tree are ordered in a natural way from "left" to "right"), then we have

$$F_{\mathcal{A}(\mathcal{B})}(x) = F_{\mathcal{A}}(F_{\mathcal{B}}(x)). \tag{3.4}$$

However, this equation is not true in general. The general formula comes out of Redfield–Pólya theory (see (4.6) and (4.9)) and requires the notion of *cycle index series*. For example, if $\mathcal{B}$ is the set of connected (unlabelled) graphs, $\mathcal{A}$ is `Sets`, so that $\mathcal{A}(\mathcal{B})$ is the set of *all* (connected and disconnected) graphs, then (3.4) is not true, but what is true is

$$F_{\mathtt{Sets}(\mathcal{B})} = \exp\left(F_{\mathcal{B}}(x) + \frac{1}{2}F_{\mathcal{B}}(x^2) + \frac{1}{3}F_{\mathcal{B}}(x^3) + \cdots\right). \tag{3.5}$$

This holds, in fact, for any set $\mathcal{B}$ of unlabelled objects. (This is seen by combining (4.9), (4.2), and (4.6).)

Next we deal with the enumeration of *labelled* objects. Let $\mathcal{A}$ be a set of labelled objects, again, each object a with a certain size $|a|$ which is a non-negative integer. "Labelled" means that each object of size n, by its structure, comes with n atoms (nodes) which are labelled $1, 2, \dots, n$. For example, $\mathcal{A}$ may be the set of all labelled graphs, where the size of a graph is the number of its vertices, and where the vertices are labelled $1, 2, \dots, n$. Again, we assume that there is only a finite number of objects from $\mathcal{A}$ of a given size. Let a_n be the number of objects from $\mathcal{A}$ of size n. The *exponential generating function* for $\mathcal{A}$ is the *formal power series*

$$E_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} \frac{x^{|a|}}{|a|!} = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!}.$$

Typical examples are `Sets` (the collection containing all "labelled sets", that is all objects of the form $\{1, 2, \dots, n\}$, including the empty set), `Permutations`, `Cycles` (labelled cycles), with respective generating functions

$$E_{\texttt{Sets}}(x) = \exp(x), \tag{3.6}$$

$$E_{\texttt{Permutations}}(x) = \frac{1}{1-x}, \tag{3.7}$$

$$E_{\texttt{Cycles}}(x) = \log \frac{1}{1-x}, \tag{3.8}$$

or `Trees` (labelled trees). The explicit form of the generating function for `Trees` is discussed in Section 5.

If $\mathcal{A}$ and $\mathcal{B}$ are two sets of objects, one defines again several other sets of objects using them. The *union* of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A} \cup \mathcal{B}$, has as a groundset the *disjoint* union of $\mathcal{A}$ and $\mathcal{B}$, and the size of an element from $\mathcal{A}$ is its size in $\mathcal{A}$, while the size of an element from $\mathcal{B}$ is its size in $\mathcal{B}$. We have

$$E_{\mathcal{A} \cup \mathcal{B}}(x) = E_{\mathcal{A}}(x) + E_{\mathcal{B}}(x). \tag{3.9}$$

To define the *product* of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A} \times \mathcal{B}$, we cannot simply take $\mathcal{A} \times \mathcal{B}$ as a groundset, we must also say something about the labelling of the objects. So, as a groundset we take all pairs (a, b) with $a \in \mathcal{A}$ and $b \in \mathcal{B}$, but labelled in all possible ways by $1, 2, \dots, |a| + |b|$ such that the order of labels assigned to a respects the original order of labels of a, and the same for b. The size of such an element (a, b) is again the sum of the sizes of a (in $\mathcal{A}$) and of b (in $\mathcal{B}$). We have

$$E_{\mathcal{A} \times \mathcal{B}}(x) = E_{\mathcal{A}}(x) \cdot E_{\mathcal{B}}(x). \tag{3.10}$$

Since, in the labelled world, objects come automatically with atoms, the substitution of two sets $\mathcal{A}$ and $\mathcal{B}$ of objects can now always be defined. The *substitution of* $\mathcal{B}$ *in* $\mathcal{A}$, denoted by $\mathcal{A}(\mathcal{B})$, is the set of objects which arises by replacing the atoms of objects from $\mathcal{A}$ by objects from $\mathcal{B}$ in all possible ways, and labelling the substituted objects in all possible ways by $1, 2, \dots, \sum_b |b|$ (the sum being over the objects from $\mathcal{B}$ which were put in the places of the atoms) that are consistent with the original labellings of the objects from $\mathcal{B}$. The size of an object from $\mathcal{A}(\mathcal{B})$ is the sum of the sizes of the objects from $\mathcal{B}$ that it contains. In order that $\mathcal{A}(\mathcal{B})$ contains only a finite number of objects of a given size, we must assume that $\mathcal{B}$ contains no elements of size 0. Then we have

$$E_{\mathcal{A}(\mathcal{B})}(x) = E_{\mathcal{A}}(E_{\mathcal{B}}(x)). \tag{3.11}$$

An example of a composition is

$$\texttt{Permutations} = \texttt{Sets}(\texttt{Cycles}).$$

Thus, from (3.11) we have

$$E_{\texttt{Permutations}}(x) = E_{\texttt{Sets}}(E_{\texttt{Cycles}}(x)),$$

corresponding to the identity

$$\frac{1}{1-x} = \exp\left(\log 1/(1-x)\right).$$

Another manifestation of the composition rule is for example the fact (which is sometimes called the "exponential principle") that, if one takes the log of the partition function for some maps, the result is the partition function for the connected maps among them.

All of the above can be generalized to a weighted setting. Namely, if $\mathcal{A}$ is a set of objects (labelled or unlabelled), and if $w : \mathcal{A} \to R$ is a *weight function* from $\mathcal{A}$ into some ring R, then all of the above remains true, if we replace the definitions of $F_{\mathcal{A}}(x)$ and $E_{\mathcal{A}}(x)$ above by the weighted sums

$$F_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} w(a)\, x^{|a|}$$

and

$$E_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} w(a) \frac{x^{|a|}}{|a|!},$$

respectively, if in the definition of the union of $\mathcal{A}$ and $\mathcal{B}$ we define the weight of an object to be its weight in $\mathcal{A}$, respectively $\mathcal{B}$, if in the definition of the product of $\mathcal{A}$ and $\mathcal{B}$ we define the weight of an object (a, b) to be the product of the weights of a and b, and if in the definition of the substitution we define the weight of an object in $\mathcal{A}(\mathcal{B})$ as the product of the weights of the objects from $\mathcal{B}$ that were put in place of the atoms.

4 Redfield–Pólya theory of coloured enumeration

The natural and uniform environment for the separate treatment of generating functions for unlabelled and labelled objects in Section 3 is the theory for counting coloured objects founded by Redfield and Pólya, in the modern treatment through cycle index series due to Joyal. We refer the reader to [4, App. 1], [7] and [40, Ch. 7] for further reading.

Let $\mathcal{A}$ be a set of labelled objects with the constraint that there is only a finite number of objects of a given size. The *cycle index series* for $\mathcal{A}$ is the formal multivariable series

$$Z_{\mathcal{A}}(x_1, x_2, \dots) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \operatorname{fix}_\sigma(\mathcal{A})\, x_1^{c_1(\sigma)} x_2^{c_2(\sigma)} x_3^{c_3(\sigma)} \cdots , \tag{4.1}$$

where $\operatorname{fix}_\sigma(\mathcal{A})$ is the number of objects a from $\mathcal{A}$ which remain invariant when the labels are permuted according to the permutation σ (in particular, if $\sigma \in \mathfrak{S}_n$, the size of a must be n in order that σ can be applied to the labels), and where $c_i(\sigma)$ denotes the number of cycles of length i of σ.

In most cases it is difficult to obtain compact expressions for the cycle index series. However, for our familiar families of objects, compact expressions are available:

$$Z_{\texttt{Sets}}(x_1, x_2, \dots) = \exp\left(x_1 + \frac{x_2}{2} + \frac{x_3}{3} + \cdots\right), \tag{4.2}$$

$$Z_{\texttt{Permutations}}(x_1, x_2, \dots) = \prod_{i=1}^{\infty} \frac{1}{1 - x_i}, \tag{4.3}$$

$$Z_{\texttt{Cycles}}(x_1, x_2, \dots) = \sum_{i=1}^{\infty} \frac{\Phi(i)}{i} \log \frac{1}{1 - x_i}, \tag{4.4}$$

where $\Phi(i)$ is the *Euler totient function* (the number of positive integers $j \le i$ relatively prime to i).

What makes the cycle index series so fundamental is the fact that the generating functions from Section 3 are specializations of it. Namely, the exponential generating function for $\mathcal{A}$ is equal to

$$E_{\mathcal{A}}(x) = Z_{\mathcal{A}}(x, 0, 0, \dots). \tag{4.5}$$

If, given the set of labelled objects $\mathcal{A}$, we produce a set of unlabelled objects $\tilde{\mathcal{A}}$ by taking all the objects from $\mathcal{A}$ but forgetting the labels, then the ordinary generating function for $\tilde{\mathcal{A}}$ is another specialization of the cycle index series,

$$F_{\tilde{\mathcal{A}}}(x) = Z_{\mathcal{A}}(x, x^2, x^3, \dots). \tag{4.6}$$

The cycle index series satisfies the following properties with respect to union, product and composition of sets of objects:

$$Z_{\mathcal{A}\cup\mathcal{B}}(x_1, x_2, \dots) = Z_{\mathcal{A}}(x_1, x_2, \dots) + Z_{\mathcal{B}}(x_1, x_2, \dots), \tag{4.7}$$

$$Z_{\mathcal{A}\times\mathcal{B}}(x_1, x_2, \dots) = Z_{\mathcal{A}}(x_1, x_2, \dots) \cdot Z_{\mathcal{B}}(x_1, x_2, \dots), \tag{4.8}$$

$$\begin{aligned} Z_{\mathcal{A}(\mathcal{B})}(x_1, x_2, \dots) = Z_{\mathcal{A}}(&Z_{\mathcal{B}}(x_1, x_2, x_3, \dots), Z_{\mathcal{B}}(x_2, x_4, x_6 \dots), \\ &Z_{\mathcal{B}}(x_3, x_6, x_9, \dots), \dots). \end{aligned} \tag{4.9}$$

Similar to the theory of generating functions surveyed in Section 3, one can also develop a weighted version of the cycle index series. Given a set of labelled objects $\mathcal{A}$, where each object a is assigned a weight $w(a)$, one changes the definition (4.1) insofar as $\mathrm{fix}_\sigma(\mathcal{A})$ gets replaced by the weighted sum $\sum_{\sigma(a)=a} w(a)$, where $\sigma(a)$ means the object arising from a by permuting the labels according to σ. Then all the above formulas remain true in this weighted setting.

Cycle index series are instrumental in the enumeration of *coloured* objects. The basic situation is that we have given a set $\tilde{\mathcal{A}}$ of unlabelled objects so that every object of size n comes with n atoms (nodes). For example, we may think of $\tilde{\mathcal{A}}$ as the set of cycles. We are now going to colour each atom by a colour from the set of colours C. The question that we pose is: *how many different coloured objects of a given size are there?* In our example, if C consists of the two colours "black" and "white", then we are asking the question of how many necklaces one can make out of n pearls that can be black or white. In terms of generating functions, we want to compute

$$\Gamma_{\tilde{\mathcal{A}}}(x) = \sum_{c} x^{|c|},$$

where the sum is over all coloured objects c that one can obtain by colouring the objects from $\tilde{\mathcal{A}}$.

The central result of Redfield–Pólya theory is that, if $\mathcal{A}$ is the set of labelled objects that one obtains from $\tilde{\mathcal{A}}$ by labelling the objects of $\tilde{\mathcal{A}}$ in all possible ways, then

$$\Gamma_{\tilde{\mathcal{A}}}(x) = Z_{\mathcal{A}}(|C|x, |C|x^2, |C|x^3, \dots).$$

There is again a weighted version. One allows the objects a from $\tilde{\mathcal{A}}$ to have weight $w(a) \in R$. Moreover, one assumes a weight function $f : C \to R$

on the colours with values in the ring R. One defines the weight of a coloured object obtained by colouring the atoms of a to be $w(a)$ multiplied by the product of all $f(\gamma)$, where γ ranges over all the colours of the atoms (including repetitions of colours). Let $\Gamma_{\tilde{\mathcal{A}}}(w, f)$ denote the sum of all the weights of all coloured objects obtained from $\tilde{\mathcal{A}}$. Then

$$\Gamma_{\tilde{\mathcal{A}}}(w, f) = Z_{\mathcal{A}}\left(\sum_{c\in C} f(c), \sum_{c\in C} f(c)^2, \sum_{c\in C} f(c)^3, \dots\right).$$

We remark that these results cover also the case of enumeration of objects under a group action. This includes the enumeration of objects on which we impose certain symmetries. See [4, App. 1], [7] and [40, Ch. 7] for more details. The enumeration of asymmetric objects is the subject of an ongoing research program (cf. [28]).

5 Solving equations for generating functions: the Lagrange inversion formula and the kernel method

In this section, we describe two methods to solve functional equations for generating functions. The Lagrange inversion makes it possible (in some situations) to find explicit expressions for the coefficients of an implicitly given series. The kernel method (and its extensions), on the other hand, is a powerful method to obtain an explicit expression for an implicitly given function. We refer the reader to [12, Sec. VII.5] for further reading.

In many situations it will happen that, when we apply the methods from Sections 3, we end up with a functional equation for the generating function $f(x) = \sum_{n=0}^{\infty} f_n x^n$ that we wanted to compute. For example, if t_n denotes the number of labelled rooted trees with n nodes, and if we write $T(x) = \sum_{n=1}^{\infty} t_n x^n/n!$, then, by applying a straight-forward decomposition of a tree into its root and its set of subtrees attached to the root, we obtain the equation

$$T(x) = x\exp(T(x)). \tag{5.1}$$

How does one solve such an equation? As a matter of fact, for $T(x)$, there is no expression in terms of known functions. However, the *Lagrange inversion formula* enables one to find the *coefficients* $t_n/n!$ of $T(x)$ explicitly. The theorem reads as follows.

Let $g(x)$ be a formal Laurent series containing only a finite number of negative powers of x, and let $f(x)$ be a formal power series without constant term. If we expand $g(x)$ in powers of $f(x)$,

$$g(x) = \sum_k c_k f^k(x), \tag{5.2}$$

then the coeffizients c_n are given by

$$c_n = \frac{1}{n}[x^{-1}]g'(x)f^{-n}(x) \quad \text{for } n \neq 0, \tag{5.3}$$

or, alternatively, by

$$c_n = [x^{-1}]g(x)f'(x)f^{-n-1}(x). \tag{5.4}$$

Here, $[x^n]h(x)$ denotes the coefficient of x^n in the power series h(x).

With this theorem in hand, the equation (5.1) is easy to solve. We write it in the form,

$$T(x)\exp(-T(x)) = x. \tag{5.5}$$

We want to know the coefficients in the expansion $T(x) = \sum_{n=0}^{\infty} t_n\, x^n/n!$. Since, by (5.5), $T(x)$ is the compositional inverse of $x\exp(-x)$, substitution of $x\exp(-x)$ instead of x gives

$$x = \sum_{n=0}^{\infty} \frac{t_n}{n!}\,(x\exp(-x))^n.$$

This equation is in the form (5.2) with $f(x) = x\exp(-x)$ and $g(x) = x$. Hence, by (5.3), we obtain

$$\frac{t_n}{n!} = \frac{1}{n}[x^{-1}](x\exp(-x))^{-n} = \frac{1}{n}[x^{n-1}]\exp(nx) = \frac{n^{n-1}}{n!},$$

and, thus, $t_n = n^{n-1}$.

The second method to solve functional equations which we explain in this section is the *kernel method*. We illustrate the method by an example. Let us consider the problem of counting Dyck paths of length $2n$ (see Section 2). Rather than attempting to arrive at a solution of the problem directly, we consider the more general problem of counting the number $a_{n,k}$ of paths consisting of steps $(1,1)$ and $(1,-1)$, which start at the origin, never drop below $y=0$, have length n and end at height k. We then form the bivariate generating function $F(u,x) = \sum_{n,k\geq 0} a_{n,k} x^n u^k$. We then have the functional equation

$$F(u,x) = 1 + xuF(u,x) + \frac{x}{u}(F(u,x) - F(0,x)), \tag{5.6}$$

since a path can be empty (this explains the term 1), it can end by a step $(1,1)$ (this explains the term $xuF(u)$), or it can end by a step $(1,-1)$. The latter can only happen if the path before that last step did not end at height 0. The generating function for these paths is $F(u,x) - F(0,x)$, and this explains the third term in the equation (5.6). In fact, we may replace (5.6) by

$$F(u,x) = 1 + xuF(u,x) + \frac{x}{u}(F(u,x) - F_1(x)), \tag{5.7}$$

because (5.7) *implies* that $F_1(x) = F(0,x)$.

The idea of the kernel method is to get rid of the unknown series $F(u,x)$. This is possible because $F(u,x)$ occurs linearly in (5.7), which can be rewritten as

$$F(u,x)\left(1 - xu - \frac{x}{u}\right) = 1 - \frac{x}{u}F_1(x). \tag{5.8}$$

We simply equate the coefficient of $F(u,x)$ in this equation to zero,

$$1 - xu - \frac{x}{u} = 0,$$

solve this for u,

$$u = \frac{1 - \sqrt{1-4x^2}}{2x},$$

(the other solution for u makes no sense in (5.7)) and substitute this back in (5.8), to obtain

$$F_1(x) = \frac{1 - \sqrt{1-4x^2}}{2x^2},$$

the familiar generating function (2.1) for the Catalan numbers. Now, by substituting this result in (5.7), we can even compute the full series $F(u,x)$.

While this was certainly a complicated, and unusual, way to compute the Catalan numbers, this approach generalizes when one considers paths with different step sets (see [12, Sec. VII.5]). In a more general situation, one has a functional equation

$$P(F(u,x), F_1(x), \ldots, F_k(x), x, u) = 0, \tag{5.9}$$

where $F(u,x)$ appears linearly, as well as the unknown series $F_1(x)$, ..., $F_k(x)$, whereas x and u appear rationally. It is clear that one can apply the same technique, namely collecting all the terms involving $F(u,x)$, equating the coefficient of $F(u,x)$ to zero, solving for u and substituting back in (5.9). If there is more than one function $F_i(x)$, then this will only give one equation for the $F_i(x)$. However, when equating the coefficient of $F(u,x)$, which was a polynomial equation, there can be more solutions. (That was actually also the case in our example, although only one solution could be used.) All these solutions can be substituted in (5.9) to give many more equations for the $F_i(x)$. The kernel method will work if we have enough equations to determine the unknown functions $F_i(x)$. See [12, Sec. VII.5] for further details. In the variant of the "obstinate kernel method", more equations are produced in more sophisticated ways. The method has been largely extended by Bousquet–Mélou and co-workers to cover equations of the form (5.9) where P is a *polynomial* such that equation (5.9) determines all involved series uniquely. This extension covers in particular the so-called *quadratic method* due to Brown, which is of great significance in the work of Tutte on the enumeration of maps. We refer the reader to [5] and the references given there for these extensions.

6 Extracting asymptotic information from generating functions

There is powerful machinery available to extract the asymptotic behaviour of the coefficients of a power series out of analytic properties of the power series. We describe the corresponding methods, *singularity analysis* and the *saddle point method* in this section. The survey [31] and the book [12] are excellent sources for further reading, which, in particular, contain several other methods which we cannot cover here for reasons of limited space.

Let us suppose that we are interested in the asymptotic behaviour of the sequence $(f_n)_{n\ge 0}$ of real (or complex) numbers as n tends to infinity. Let us suppose that the power series $f(z)=\sum_{n=0}^{\infty} f_n\, z^n$ converges in some neighbourhood of the origin. (If this series converges only at $z=0$, then either one has to try to scale, that is, for example, look at the power series $f(z)=\sum_{n=0}^{\infty} f_n\, z^n/n!$ instead, or one must apply methods other than singularity analysis or the saddle point method. In the latter case, depending on the nature of the coefficients f_n, this may be the *Euler–Maclaurin* or the *Poisson summation formulas*, the *Mellin transform technique*, or other direct methods. The reader is referred to [12, 31].) The idea is then to consider $f(z)$ as a complex function in z (and extend the range of f beyond the disk of convergence about the origin), and to study the singularities of $f(z)$. (The point at infinity can also be a singularity.) The upshot is that the singularities of $f(z)$ with smallest modulus dictate the asymptotic behaviour of the coefficients f_n. These singularities of smallest modulus are called the *dominating* singularities.

If there is an infinite number of dominant singularities, then one has to try the *circle method*. We refer the reader to [1, 3] for details of this method.

If there is a finite number of dominant singularities, then there can be again two different situations, depending on whether these are "small" or "large" singularities. Roughly speaking, a singularity is small if the function $f(z)$ grows at most polynomially when z approaches the singularity, otherwise it is "large". A typical example of a small singularity is $z = 1/4$ in $(1 - 4z)^{-1/2}$, whereas a typical example of a large singularity is $z = \infty$ in $\exp(x)$ or $z = 1$ in $\exp(1/(1 - z))$.

The method to apply for small singularities is the method of *singularity analysis* as developed by Flajolet and Odlyzko. (Singularity analysis implies *Darboux's method*, which occurs frequently in the literature, and, thus, supersedes it.) For the sake of simplicity, we consider first only the case of a unique dominant singularity. We shall address the issue of several dominant singularities a little bit later. Furthermore, we assume the singularity to be $z = 1$, again for the sake of simplicity of presentation. The general result can then be obtained by rescaling z.

The basic idea is the transfer principle

$$\text{If } f(z) \underset{z\to 1}{=} \sigma(z) + O(\tau(z)) \quad \text{then} \quad f_n \underset{n\to\infty}{=} \sigma_n + O(\tau_n), \tag{6.1}$$

where $\sigma(z) = \sum_{n=0}^{\infty} \sigma_n z^n$ is a linear combination of standard functions of the form $(1-z)^{-\alpha}$, or logarithmic variants, and $\tau(z) = \sum_{n=0}^{\infty} \tau_n z^n$ also lies in the scale. (See [12, Sec. VI.3,4] for the exact statement.) The expansion for $f(z)$ in (6.1) is called the *singular expansion* of $f(z)$. For the above mentioned standard functions, we have

$$\begin{aligned}[z^n](1-z)^{-\alpha}&\left(\frac{1}{z}\log\frac{1}{1-z}\right)^{\beta}\\ &\sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{\beta}\left(1+\frac{C_1}{1!}\frac{\beta}{\log n}+\frac{C_2}{2!}\frac{\beta(\beta-1)}{(\log n)^2}+\cdots\right),\end{aligned} \tag{6.2}$$

where $[z^n]g(z)$ denotes the coefficient of z^n in $g(z)$, and where

$$C_k = \Gamma(\alpha)\frac{d^k}{ds^k}\frac{1}{\Gamma(s)}\bigg|_{s=\alpha}.$$

If α is a non-positive integer, then this expansion has to be taken with care (cf. [12, Sec. VI.2]).

To see how this works, consider the example $f_n = \sum_{k=0}^{n} \binom{2k}{k}$. We have

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

The function on the right-hand side is meromorphic in all of $\mathbb{C}$ (where $\mathbb{C}$ denotes the complex numbers), with singularities at $z = 1$ and $z = 1/4$. The dominant singularity is $z = 1/4$. We determine the singular expansion of $f(z)$ about $z = 1/4$,

$$f(z) = \frac{4}{3}(1-4z)^{-1/2} - \frac{4}{9}(1-4z)^{1/2} + \frac{4}{27}(1-4z)^{3/2} + O\left((1-4z)^{5/2}\right).$$

(We stopped the expansion after 3 terms. The farther we go, the more terms can we compute of the asymptotic expansion for f_n.) Hence, we obtain

$$\begin{aligned} f_n &= 4^n \Bigg(\frac{4}{3} \frac{n^{-1/2}}{\Gamma(1/2)} \left(1 - \frac{1}{8n} + \frac{1}{128n^2}\right) - \frac{4}{9} \frac{n^{-3/2}}{\Gamma(-1/2)} \left(1 + \frac{3}{8n}\right) \\ &\qquad\qquad + \frac{4}{27} \frac{n^{-5/2}}{\Gamma(-3/2)} + O\left(n^{-7/2}\right) \Bigg) \\ &= \frac{4^n}{\sqrt{\pi n}} \left(\frac{4}{3} + \frac{1}{18n} + \frac{11}{288n^2} + O\left(\frac{1}{n^3}\right)\right). \end{aligned}$$

If there is several small dominant singularities (but only a finite number of them), then one simply applies the above procedure for all of them and, to obtain the desired asymptotic expansion, one adds up the corresponding contributions.

The method to apply for large singularities is the *saddle point method.* For the following considerations we assume that $f(z)$ is analytic in $|z| < R \le \infty$. At the heart of the saddle point method lies Cauchy's formula

$$f_n = [z^n]f(z) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f(z)}{z^{n+1}} dz \tag{6.3}$$

for writing the n-th coefficient in the power series expansion of $f(z)$. Here, $\mathcal{C}$ is some simple closed contour around the origin that stays in the range $|z| < R$. The idea is to exploit the fact that we are free to deform the contour. The aim is to choose a contour such that the main contribution to the integral in (6.3) comes from a very tiny part of the contour, whereas the contribution of the rest is negligible. This will be possible if we put the contour through a saddle point of the integrand $f(z)/z^{n+1}$. Under suitable conditions, the main contribution will then come from the small passage of the path through the saddle point, and the contribution of the rest will be negligible.

In practice, the the saddle point method is not always straight-forward to apply, but has to be adapted to the specific properties of the function $f(z)$ that we are encountering. I refer the reader to the corresponding chapters in [12, 31] for more details. There is one important exception though, namely the *Hayman admissible functions*. I shall not reproduce the definition of Hayman admissibility because it is cumbersome (cf. [12, Sec. VIII.5] and [31, Def. 12.4]). However, in many applications it is not even necessary to go back to it because of the closure properties of Hayman admissible functions. Namely, it is known (cf. [31, Theorem 12.8]) that $\exp(p(z))$ is Hayman admissible in $|z| < \infty$ for any polynomial $p(z)$ with real coefficients as long as the coefficients a_n of the Taylor series of $\exp(p(z))$ are positive for all sufficiently large n (thus, $\exp(z)$ is Hayman admissible, for example), and it is known that, if $f(z)$ and $g(z)$ are Hayman admissible in $|z| < R \le \infty$, then $\exp(f(z))$ and $f(z)g(z)$ are also (thus, $\exp(\exp(z) - 1)$ is Hayman admissible, for example).

The central result of Hayman's theory is the following: *If $f(z) = \sum_{n\ge 0} f_n z^n$ is Hayman admissible in $|z| < R$, then*

$$f_n \sim \frac{f(r_n)}{r_n^n \sqrt{2\pi b(r_n)}} \quad \text{as } n \to \infty, \tag{6.4}$$

where r_n is the unique solution for large n of the equation $a(r) = n$ in (R_0, R), with $a(r) = rf'(r)/f(r)$, $b(r) = ra'(r)$, and a suitably chosen constant $R_0 > 0$.

This result covers only the first term in the asymptotic expansion. There is an even more sophisticated theory due to Harris and Schoenfeld which allows one to also find a complete asymptotic expansion. We refer the reader to [12, Sec. VIII.5] and [31, pp. 1180ff] for more details.

Methods for the asymptotic analysis of multivariable generating functions are also available, see the corresponding chapters in [12, 31], and the recent important development surveyed in [32]. We add that both the method of singularity analysis and Hayman's theory of admissible functions have been made largely automatic, and that this has been implemented in the *Maple*

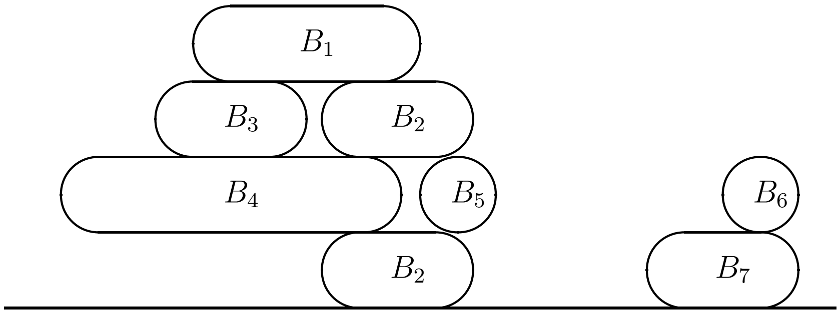


Figure 6: A heap of pieces

program `gdev`.[1]

7 The theory of heaps

The *theory of heaps*, developed by Viennot, is a geometric rendering of the theory of the partial commutation monoid of Cartier and Foata, which is now most often called the *Cartier–Foata monoid.* Its importance stems from the fact that several objects which appear in statistical physics, such as Motzkin paths, animals, respectively polyominoes, or Lorentzian triangulations (see [43] and the references therein), are in bijection with heaps.

Informally, a *heap* is what we would imagine. We take a collection of "pieces," say $B_1, B_2, \dots$, and put them one upon the other, sometimes also sideways, to form a "heap," see Figure 6.

There, we imagine that pieces can only move vertically, so that the heap in Figure 6 would indeed form a stable arrangement. Note that we allow several copies of a piece to appear in a heap. (This means that they differ only by a vertical translation.) For example, in Figure 6 there appear two copies of B_2. Under these assumptions, there are pieces which can move past each other, and others which cannot. For example, in Figure 6, we can move the piece B_6 higher up, thus moving it higher than B_1 if we wish. However, we cannot move B_7 higher than B_6, because B_6 blocks the way. On the other hand, we can move B_7 past B_1 (thus taking B_6 with us). Thus, a rigorous way to introduce heaps is by beginning with a set $\mathcal{B}$ of pieces (in our example, $\mathcal{B} = \{B_1, B_2, \dots, B_7\}$), and we declare which pieces can be moved

[1] The program is available at `http://algo.inria.fr/libraries/libraries.html#gdev`.

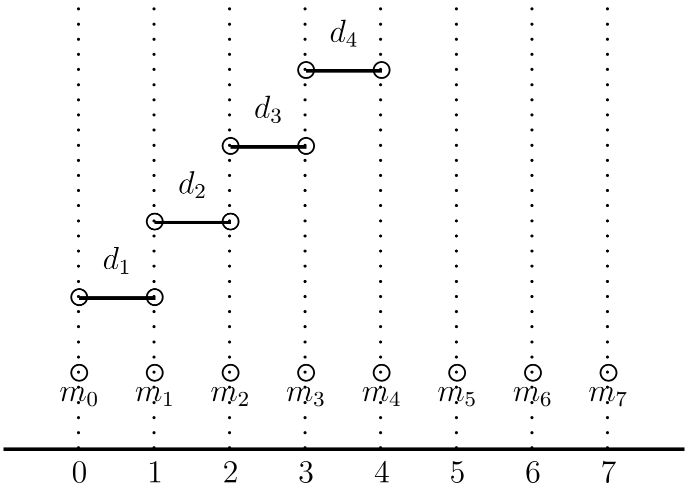


Figure 7: Monomers and dimers

past another and which cannot. We indicate this by a symmetric relation $\mathcal{R}$: we write $a\mathcal{R}b$ to indicate that a *cannot* move past b (and vice versa). When we consider a word $a_1a_2\dots a_n$ of pieces, $a_i \in \mathcal{B}$, we think of it as putting first a_1, then putting a_2 on top of it (and, possibly, moving it past a_1), then putting a_3 on top of what we already have, etc. We declare two words to be equivalent if one arises from the other by commuting adjacent letters which are not in relation. A heap is then an equivalence class of words under this equivalence relation. What we have described just now is indeed the original definition of Cartier and Foata.

The class of heaps which occurs most frequently in applications is the class of heaps of *monomers* and *dimers*, which we now introduce. Let $\mathcal{B} = M \cup D$, where $\mathcal{M} = \{m_0, m_1, \dots\}$ is the set of monomers and $\mathcal{D} = \{d_1, d_2, \dots\}$ is the set of dimers. We think of a monomer m_i as a point, symbolized by a circle, with x-coordinate i, see Figure 7. We think of a dimer d_i as two points, symbolized by circles, with x-coordinates $i-1$ and i which are connected by an edge, see Figure 7. We impose the relations $m_i\mathcal{R}m_i$, $m_i\mathcal{R}d_i$, $m_i\mathcal{R}d_{i+1}$, $i = 0, 1, \dots$, $d_i\mathcal{R}d_j$, $i-1 \le j \le i$, and extend $\mathcal{R}$ to a symmetric relation. Figure 8 shows two heaps of momomers and dimers.

For example, Motzkin paths are in bijection with heaps of monomers and dimers. To see this, given a Motzkin path, we read the steps of the path from the beginning to the end. Whenever we read a level-step at height h, we make it into a monomer with x-coordinate h, whenever we read a down-step from height h to height $h-1$, we make it into a dimer whose end points have

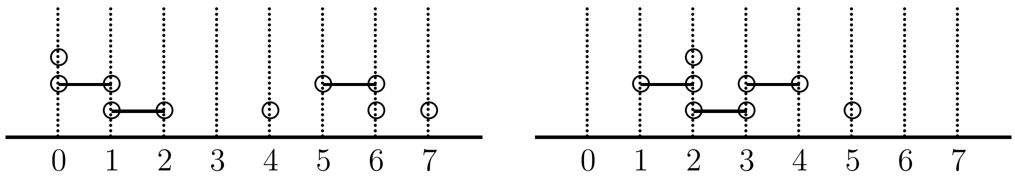


Figure 8: Two heaps of monomers and dimers

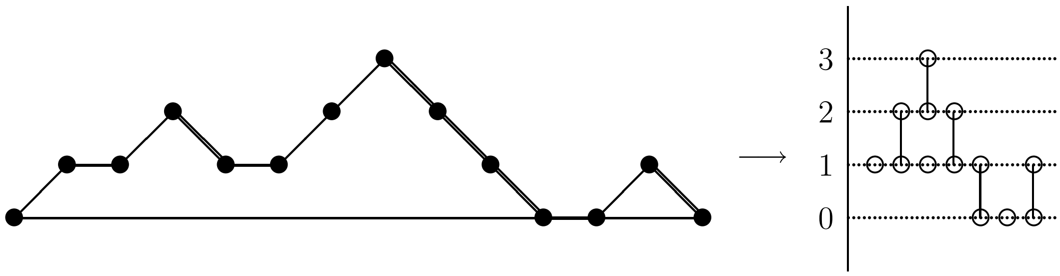


Figure 9: Bijection between Motzkin paths and heaps of monomers and dimers

x-coordinates $h-1$ and h. Up-steps are ignored. Figure 9 shows an example. In the figure, the heap is not in "standard" fashion, in the sense that the x-axis is not shown as a horizontal line but as a vertical line (compare with Figure 7). But it could be easily transformed into "standard" fashion by a simple reflection with respect to a line of slope 1.

Lattice animals on the triangular lattice and on the quadratic lattice are also in bijection with heaps, this time with heaps consisting entirely out of dimers. Given an animal, one simply replaces each vertex of the animal by a dimer, see Figures 10 and 11. While in the case of animals on the triangular lattice this gives a constraint-less bijection (see Figure 10), in the case of the quadratic lattice this sets up a bijection with heaps of dimers *in which two dimers of the same type can never be placed directly one over the other* (see Figure 11). For example, two dimers d_5, one placed directly over the other (as they occur in Figure 10), are forbidden under this rule.

Next we make heaps into a *monoid* by introducing a *composition* of heaps. (A monoid is a set with a binary operation which is associative.) Intuitively, given two heaps H_1 and H_2, the composition of H_1 and H_2, the heap $H_1 \circ H_2$, is the heap which results by putting H_2 on top of H_1. In terms of words, the composition of two heaps is the equivalence class of the concatenation uw, where u is a word from the equivalence class of H_1, and w is a word from the

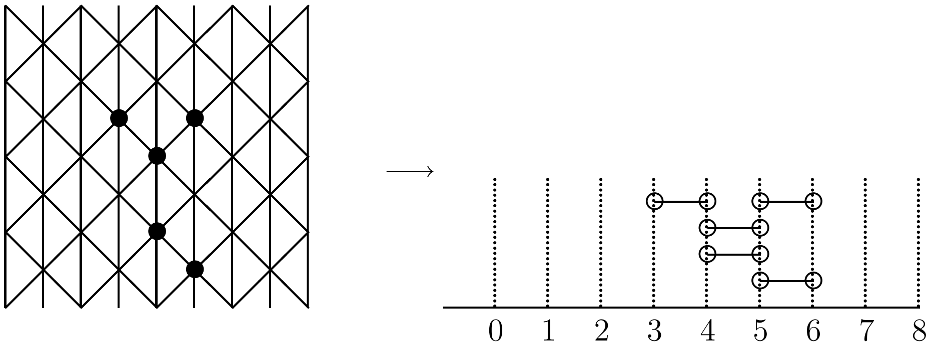


Figure 10: Bijection between animals and heaps of dimers

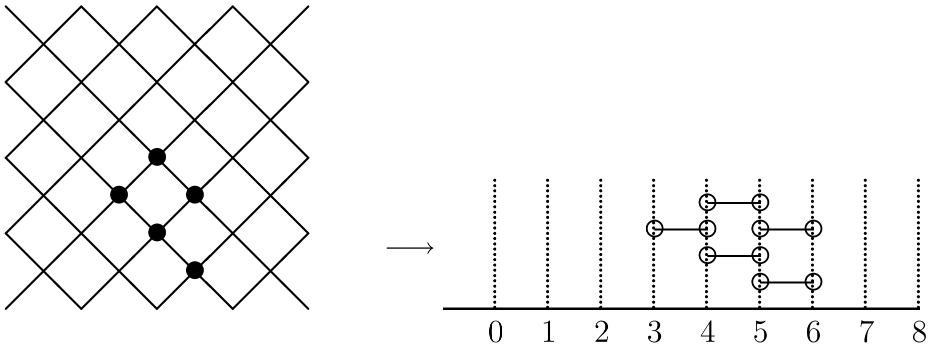


Figure 11: Bijection between animals and heaps of dimers

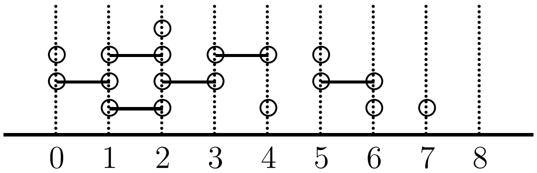


Figure 12: The composition of the heaps in Figure 8

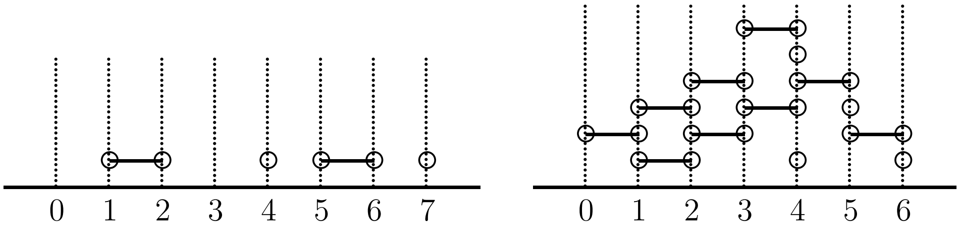


Figure 13: a. A trivial heap b. A pyramid

equivalence class of H_2.

The composition of the two heaps in Figure 8 is shown in Figure 12.

Given pieces $\mathcal{B}$ with relation $\mathcal{R}$, let $\mathcal{H}(\mathcal{B}, \mathcal{R})$ be the set of all heaps consisting of pieces from $\mathcal{B}$, including the empty heap, the latter denoted by $\emptyset$. It is easy to see that the composition makes $(\mathcal{H}(\mathcal{B}, \mathcal{R}), \circ)$ into a monoid with unit $\emptyset$.

For the statement of the main theorem in the theory of heaps, we need two more terms. A *trivial heap* is a heap consisting of pieces all of which are pairwise unrelated. Figure 13.a shows a trivial heap consisting of monomers and dimers. A *pyramid* is a heap with exactly one maximal (= top-most) element. Figure 13.a shows a pyramid consisting of monomers and dimers. Finally, if H is a heap, then we write $|H|$ for the number of pieces in H.

In applications, heaps will have weights, which are defined by introducing a weight $w(B)$ for each piece B in $\mathcal{B}$, and by extending the weight w to all heaps H by letting $w(H)$ denote the product of all weights of the pieces in H (multiplicities of pieces included).

Let $\mathcal{M}$ be a subset of the pieces $\mathcal{B}$. Then, the generating function for all heaps with maximal pieces contained in $\mathcal{M}$ is given by

$$\sum_{\substack{H \in \mathcal{H}(\mathcal{B}, \mathcal{R}) \\ \text{maximal pieces } \subseteq \mathcal{M}}} w(H) = \frac{\sum_{T \in \mathcal{T}(\mathcal{B} \setminus \mathcal{M}, \mathcal{R})} (-1)^{|T|} w(T)}{\sum_{T \in \mathcal{T}(\mathcal{B}, \mathcal{R})} (-1)^{|T|} w(T)}, \tag{7.1}$$

where $\mathcal{T}(\mathcal{B},\mathcal{R})$ denotes the set of all trivial heaps with pieces from $\mathcal{B}$. In particular, the generating function for *all* heaps is given by

$$\sum_{H\in\mathcal{H}(\mathcal{B},\mathcal{R})} w(H) = \frac{1}{\sum_{T\in\mathcal{T}(\mathcal{B},\mathcal{R})}(-1)^{|T|}w(T)}. \tag{7.2}$$

Furthermore, if $\mathcal{P}(\mathcal{B},\mathcal{R})$ denotes the set of all pyramids with pieces from $\mathcal{B}$, then

$$\sum_{P\in\mathcal{P}(\mathcal{B},\mathcal{R})} \frac{w(P)}{|P|} = \log\Bigg(\sum_{H\in\mathcal{H}(\mathcal{B},\mathcal{R})} w(H)\Bigg), \tag{7.3}$$

where $|P|$ is the number of pieces of P. (As the reader will have guessed, this is a consequence of the "exponential principle" mentioned in Section 3.)

8 The transfer matrix method

The transfer matrix method (cf. [39, Ch. 4] for further reading) applies whenever we are able to build the combinatorial objects that we are interested in by moving on a finite number of states in a step-by-step fashion, where the current step does not depend on the previous ones. (In statistical language, we are considering a *finite state Markov chain.*) For example, Motzkin paths which are constrained to stay between two parallel lines, say between $y = 0$ and $y = K$, can be described in such a way: the states are the heights $0, 1, \dots, K$, and, if we are in state h, then in the next step we are allowed to move to states $h + 1$, h, or $h - 1$, except that from state 0 we cannot move to -1 (there is no state -1), and when we are in state K we cannot move to $K + 1$ (there is no state $K + 1$).

For describing the general situation, let $G = (V, E)$ be a directed graph with vertex set V and edge set E. Let $w_n(u, v)$ denote the number of walks from vertex u to vertex v along edges of G. To compute these numbers, we consider the *adjacency matrix* of G, $A(G)$. By definition, using our notation, $A(G) = (w_1(u, v))_{u,v\in V}$. Obviously, $(w_n(u, v))_{u,v\in V} = (A(G))^n$. Thus,

$$\left(\sum_{n=0}^{\infty} w_n(u,v)\,x^n\right)_{u,v\in V} = \sum_{n=0}^{\infty}(A(G))^n\,x^n = (I_n - A(G)x)^{-1}\,,$$

where I_n is the $n\times n$ identity matrix. In other words, the generating functions $\sum_{n=0}^{\infty} w_n(u, v)\,x^n$ for the walk numbers between u and v form the entries of a matrix which is the inverse matrix of $I_n - A(G)x$. By elementary linear algebra,

$$\sum_{n=0}^{\infty} w_n(u,v)\, x^n = \frac{(-1)^{\#u+\#v} \det(I_n - A(G)x)_{v,u}}{\det(I_n - A(G)x)}, \tag{8.1}$$

where $\det(I_n - A(G)x)_{v,u}$ is the minor of $I_n - A(G)x$ with the row indexed by v and the column indexed by u omitted, and where $\#u$ denotes the row number of u and similarly for $\#v$. A weighted version could also be developed in the same way, where we put a weight $w(e)$ on each edge, and the weight of a walk is the product of the weights of all its edges.

In particular, the expression (8.1) is a *rational function* in x. Then, by the basic theorem on rational generating functions (cf. [39, Sec. 4.1]), the number $w_n(u,v)$ can be expressed as a sum $\sum_{i=1}^{d} P_i(n)\gamma_i^n$, where the γ_i's are the different roots of the polynomial $\det(xI_n - A(G))$, and $P_i(n)$ is a polynomial of degree less than the multiplicity of the root γ_i. (The $P_i(n)$'s depend on u and v, whereas the γ_i's do not.) If there exists a unique root γ_j with maximal modulus, then this implies that, asymptotically as $n \to \infty$, $w_n(u,v) \sim P_j(n)\gamma_j^n$.

9 Lattice paths

Recall from Section 2, that a *lattice path* P in $\mathbb{Z}^d$ is a path in the d-dimensional integer lattice $\mathbb{Z}^d$ which uses only points of the lattice, that is, it is a sequence $(P_0, P_1, \dots, P_l)$, where $P_i \in \mathbb{Z}^d$ for all i. The vectors $\overrightarrow{P_0P_1}, \overrightarrow{P_1P_2}, \dots, \overrightarrow{P_{l-1}P_l}$ are called the *steps* of P. The number of steps, l, is called the *length* of P.

The enumeration of lattice paths has always been an intensively studied topic in statistics, because of their importance in the study of *random walks*, of *rank order statistics* for *non-parametric testing*, and of *queueing processes*. The reader is referred to [10] and particularly Mohanty's book [30], which is a rich source for enumerative results on lattice paths, albeit in a statistical language. We review the most important results in this section. Most of these concern 2-dimensional lattice paths, i.e., the case $d = 2$.

To begin with, we consider paths in the integer plane $\mathbb{Z}^2$ consisting of horizontal and vertical unit steps in the positive direction. Clearly, the number of all (unrestricted) paths from the origin to (n,m) is the binomial coefficient $\binom{n+m}{n}$. By the *reflection principle*, which is commonly attributed to D. André (see e.g. [8, p. 22]), it follows that the number of paths from the origin to (n,m) which do not pass above the line $y = x + t$, where $m \le n + t$, is given by[2]

$$\binom{n+m}{n} - \binom{n+m}{n+t+1}. \tag{9.1}$$

If one considers more generally paths bounded by the line $my = nx + t$, no compact formula is known. It seems that the most conceptual way to approach this problem is through the so-called *kernel method* (see Section 5), which, in combination with the saddle point method, allows one also to obtain strong asymptotic results. There is one special instance, however, which has a "nice" formula. The number of all lattice paths from the origin to (n, m) which never pass above $x = \mu y$, where μ is a positive integer, is given by

$$\frac{n - \mu m + 1}{n + m + 1}\binom{n+m+1}{m}. \tag{9.2}$$

The most elegant way to prove this formula is by means of the *cycle lemma* of Dvoretzky and Motzkin (see [30, p. 9], where the cycle lemma occurs under the name of "penetrating analysis").

Iteration of the reflection principle shows that the number of paths from the origin to (n, m) which stay between the lines $y = x + t$ and $y = x + s$ (being allowed to touch them), where $t \geq 0 \geq s$ and $n + t \geq m \geq n + s$, is given by the finite (!) sum (see e.g. [30, p. 6])

$$\sum_{k\in\mathbb{Z}} \left(\binom{n+m}{n - k(t-s+2)} - \binom{n+m}{n - k(t-s+2) + t + 1} \right). \tag{9.3}$$

The enumeration of lattice paths restricted to regions bounded by hyperplanes has also been considered for other regions, such as quadrants, octants, and rectangles, as well as in higher dimensions. A general result due to Gessel and Zeilberger, and Biane, independently, on the number of lattice paths in a chamber (alcove) of an (affine) reflection group (see [23] for the corresponding references and pointers to further results) shows how far one can go when one uses the reflection principle. In particular, this result covers (9.1) and (9.3), the enumeration of lattice paths in quadrants, octants, rectangles,

[2] Roughly, the reflection principle sets up a bijection between the paths from the origin to (n, m) which *do pass above* the line $y = x + t$ and *all* paths from $(-t-1, t+1)$ to (n, m), by reflecting the path portion between the origin and the last touching point on $y = x+t+1$ in this latter line. Thus, the result of the enumeration problem is the number of *all* paths from $(0, 0)$ to (n, m), which is given by the binomial coefficient $\binom{n+m}{n}$, minus the number of *all* paths from $(-t-1, t+1)$ to (n, m), which is given by the binomial coefficient $\binom{n+m}{n+t+1}$, whence the formula (9.1).

and many other results that have appeared (before and after) in the literature. We present a particularly elegant (and frequently occurring) special case. (In reflection group language, it corresponds to the reflection group of "type A_{n-1}". See [18] for terminology and information on reflection groups.)

Let $A = (a_1, a_2, \dots, a_d)$ and $E = (e_1, e_2, \dots, e_d)$ be points in $\mathbb{Z}^d$ with $a_1 \geq a_2 \geq \cdots \geq a_d$ and $e_1 \geq e_2 \geq \cdots \geq e_d$. The number of all paths from A to E in the integer lattice $\mathbb{Z}^d$, which consist of positive unit steps and which stay in the region $x_1 \geq x_2 \geq \cdots \geq x_d$, equals

$$\Big(\sum_{i=1}^{d}(e_i - a_i)\Big)! \det_{1\leq i,j\leq d}\left(\frac{1}{(e_i - a_j - i + j)!}\right). \tag{9.4}$$

The counting problem of the theorem is equivalent to numerous other counting problems. It has been originally formulated as an n-candidate ballot problem, but it is as well equivalent to counting the number of standard Young tableaux of a given shape. In the case that all a_j's are equal, the determinant does in fact evaluate into a closed form product. In Young tableaux theory a particular way to write the result is known as the *hook-length formula* (see e.g. [40, Cor. 7.21.6]).

We return to lattice paths in the plane, mentioning some more closely related results. The first is a result of Mohanty [30, Sec. 4.2], which expresses the number of all lattice paths from the origin to (n, m) which touch the line $y = x + t$ *exactly* r times, never crossing it, as the difference

$$\binom{n+m-r}{n+t-1} - \binom{n+m-r}{n+t}, \qquad r \geq 1. \tag{9.5}$$

Not forbidding that the paths cross the bounding line, we arrive at the problem of counting the lattice paths from the origin to (n, m), which cross the main diagonal $y = x$ exactly r times, the answer being

$$\begin{cases} \frac{m-n+2r+1}{m+n+1}\binom{m+n+1}{n-r} & \text{if } m > n, \\ \frac{2r+2}{n}\binom{2n}{n-r-1} & \text{if } m = n. \end{cases} \tag{9.6}$$

Next, we give the number of lattice paths from the origin to (n, n) which have $2r$ steps on one side of the line $y = x$, as

$$\binom{2r}{r}\binom{2n-2r}{n-r}, \tag{9.7}$$

a result due to Sparre Andersen. We refer the reader to [30, Ch. 3] for further results in this direction.

Enumerating lattice paths with a fixed number of maximal straight pieces (which correspond to runs), is intimately connected to another basic enumeration problem concerning lattice paths: the enumeration of lattice paths having a fixed number of *turns*. An effective way to attack the latter problem is by means of *two-rowed arrays*, see the survey article [22], where in particular analogues of the reflection principle for two-rowed arrays are developed. These imply formulas for the number of lattice paths with fixed starting and end points and a fixed number of *north-east* (respectively *east-north*) *turns*, for unrestricted paths, as well as for paths bounded by lines. (A north-east turn in a lattice path is a point where the direction changes from "north" to "east." An east-north turn is defined analogously.) In particular, analogues of (9.1)–(9.3) are known when the number of north-east (respectively east-north) turns is fixed.

These formulas imply for example (see again [22, Sec. 3.5]) that the number of lattice paths from the origin to (n,n) which never pass above the line $y=x+t$ and have exactly $2r$ maximal straight pieces is given by

$$2\binom{n-1}{r-1}^2-\binom{n+t-1}{r-2}\binom{n-t-1}{r}-\binom{n+t-1}{r-1}\binom{n-t-1}{r-1}, \quad (9.8)$$

with a similar result for the case of $2r+1$ maximal straight pieces.[3] Furthermore, they imply that the number of lattice paths from the origin to (n,n) which never pass above the line $y=x+t$ and never below the line $y=x-t$ and have exactly $2r$ maximal straight pieces is given by

$$\sum_{k=-\infty}^{\infty}\left\{2\binom{n-2kt-1}{r+k-1}\binom{n+2kt-1}{r-k-1}\right.\\ -\binom{n-2kt+t-1}{r+k-2}\binom{n+2kt-t-1}{r-k}\\ \left.-\binom{n-2kt+t-1}{r+k-1}\binom{n+2kt-t-1}{r-k-1}\right\}, \quad (9.9)$$

with a similar result for the case of $2r+1$ maximal straight pieces.

[3]If $t=0$, the numbers in (9.8) become $\frac{1}{n}\binom{n}{r}\binom{n}{r-1}$, and they are known as the *Narayana numbers*.

The most general boundary for lattice paths that one can imagine is the restriction that it stays between two given (fixed) paths. Let us assume that the horizontal steps of the upper (fixed) path are at heights $a_1 \le a_2 \le \cdots \le a_n$, whereas the horizontal steps of the lower (fixed) path are at heights $b_1 \le b_2 \le \cdots \le b_n$, $a_i \ge b_i$, $i = 1, 2, \dots, n$. Then the number of all paths from $(0, b_1)$ to (n, a_n) satisfying the property that for all $i = 1, 2, \dots, n$ the height of the i-th horizontal step is between b_i and a_i is given by the determinant

$$\det_{1\le i,j\le n}\left(\binom{a_i - b_j + 1}{j - i + 1}\right). \tag{9.10}$$

In the statistical literature, this formula is often known as "Steck's formula", but it is actually a special case of a much more general theorem due to Kreweras. A generalization of (9.10) to higher dimensional paths was given by Handa and Mohanty (see [30, Sec. 2.4].

Next, we consider three-step lattice paths in the integer plane $\mathbf{Z}^2$, that is, paths consisting of up-steps $(1, 1)$, level steps $(1, 0)$, and down-steps $(1, -1)$. The particular problem that we are interested in is to count such three-step paths starting at $(0, r)$ and ending at (ℓ, s), which do not pass below the x-axis and do not pass above the horizontal line $y = K$. Furthermore, we assign the weight 1 to an up-step, the weight b_h to a level-step at height h, and the weight λ_h to a down-step from height h to $h - 1$. The weight $w(P)$ of a path P is defined as the product of the weights of all its steps. Then we have the following result, which can be obtained by the transfer matrix method described in Section 8.

Define the sequence $(p_n(x))_{n\ge0}$ *of polynomials by*

$$xp_n(x) = p_{n+1}(x) + b_n p_n(x) + \lambda_n p_{n-1}(x), \qquad \text{for } n \ge 1, \tag{9.11}$$

with initial conditions $p_0(x) = 1$ *and* $p_1(x) = x - b_0$. *Furthermore, define* $(Sp_n(x))_{n\ge0}$ *to be the sequence of polynomials which arises from the sequence* $(p_n(x))$ *by replacing* λ_i *by* λ_{i+1} *and* b_i *by* b_{i+1}, $i = 0, 1, 2, \dots,$ *everywhere in the three-term recurrence* (9.11) *and in the initial conditions. Finally, given a polynomial* $p(x)$ *of degree* n, *we denote the corresponding reciprocal polynomial* $x^n p(1/x)$ *by* $p^*(x)$.

With the weight w *defined as before, the generating function* $\sum_P w(P)x^{\ell(P)}$, *where the sum is over all three-step paths* P *which start at* $(0, r)$, *terminate at height* s, *do not pass below the* x*-axis, and do not pass above the line* $y = K$, *and where $\ell(P)$ denotes the length of P, is given by*

$$\begin{cases} \dfrac{x^{s-r}p_r^*(x)S^{s+1}p_{K-s}^*(x)}{p_{K+1}^*(x)} & r \le s, \\ \lambda_r \cdots \lambda_{s+1}\dfrac{x^{r-s}p_s^*(x)S^{r+1}p_{K-r}^*(x)}{p_{K+1}^*(x)} & r \ge s. \end{cases} \tag{9.12}$$

The sequence of polynomials $(p_n(x))_{n\ge 0}$ is in fact a sequence of *orthogonal polynomials* (cf. [21, 41]).

We remark that in the case that $r = s = 0$ there is also an elegant expression for the generating function due to Flajolet (see [12, Sec. V.2]) in terms of a continued fraction.

In order to solve our problem, we just have to extract the coefficient of x^ℓ in (9.12). By a partial fraction expansion, a formula of the type

$$\sum_m c_m \xi_m^\ell, \tag{9.13}$$

results, where the ξ_m's are the zeroes of $p_{K+1}(x)$, and the c_m's are some coefficients, only a finite number of them being non-zero.

It should be noted that, because of the many available parameters (the b_n's and λ_n's), by appropriate specializations one can also obtain numerous results about enumerating three-step paths according to various statistics, such as the number of touchings on the bounding lines, etc.

There are two important special cases, in which a completely explicit solution in terms of elementary functions can be given.

The first case occurs for $b_i = 0$ and $\lambda_i = 1$ for all i. In this case, the polynomials $p_n(x)$ defined by the three-term recurrence (9.11) are Chebyshev polynomials of the second kind,[4] $p_n(x) = U_n(x/2)$. The result which is then obtained from the general theorem (clearly, the zeroes of $U_n(x)$ are $x = \cos(2k\pi/(n+1))$, $k = 1, 2, \dots, n$, and therefore the partial fraction expansion of (9.12) is easily determined) is that the number of lattice paths from $(0, r)$ to (ℓ, s) with only up- and down-steps, which always stay between the x-axis and the line $y = K$, is given by (see also [10, Ch. XIV, Eq. (5.7)])

$$\frac{2}{K+2}\sum_{k=1}^{K+1}\left(2\cos\frac{\pi k}{K+2}\right)^\ell \cdot \sin\frac{\pi k(r+1)}{K+2} \cdot \sin\frac{\pi k(s+1)}{K+2}, \tag{9.14}$$

[4]The Chebyshev polynomial of the second kind $U_n(x)$ is defined by $U_n(\cos t) = \sin((n+1)t)/\sin t$ (see [21] for almost exhaustive information on these polynomials and, more generally, on hypergeometric orthogonal polynomials).

a formula which goes back to Lagrange.

The second case occurs for $b_i = 1$ and $\lambda_i = 1$ for all i. In this case, the polynomials $p_n(x)$ defined by the three-term recurrence (9.11) are again Chebyshev polynomials of the second kind, $p_n(x) = U_n((x-1)/2)$. The result which is then obtained from the general theorem is that the number of three-step lattice paths from $(0,r)$ to (ℓ,s), which always stay between the x-axis and the line $y=K$, is given by

$$\frac{2}{K+2}\sum_{k=1}^{K+1}\left(2\cos\frac{\pi k}{K+2}+1\right)^{\ell}\cdot\sin\frac{\pi k(r+1)}{K+2}\cdot\sin\frac{\pi k(s+1)}{K+2}. \tag{9.15}$$

10 Perfect matchings and tilings

In this section we consider the problem of counting the perfect matchings of a graph. For an introduction into the problem, and into methods to solve it, as well as for a report on recent developments, we refer the reader to [34].

Let $G=(V,E)$ be a finite loopless graph with vertex set V and edge set E. A *matching* (also called 1-*factor* in graph theory) is a subset of the edges with the property that no two edges share a vertex. A matching is *perfect* if it covers all the edges. Let $M(G)$ denote the number of perfect matchings of the graph G. More generally, we could assign a weight $w(e)$ to each edge e of the graph and define the weight of a matching to be the product of the weights of all its edges. Let $M_w(G)$ denote the sum of all weights of all matchings of the graph G.

Kasteleyn's method for determining $M(G)$, respectively $M_w(G)$, makes use of determinants and Pfaffians. Recall that the *Pfaffian* $\operatorname{Pf}(A)$ of a triangular array $A=(a_{i,j})_{1\le i<j\le 2n}$ is defined by

$$\operatorname{Pf}(A)=\sum_m(\operatorname{sgn} m)\prod_{\{i,j\}\in m}a_{i,j}, \tag{10.1}$$

where the sum is over all perfect matchings of the complete graph on vertices $\{1,2,\dots,2n\}$, and where the product is over all edges $\{i,j\}$, $i<j$, of m. The sign $\operatorname{sgn} m$ of m is $(-1)^{\#\text{crossings of } m}$, where a crossing is a pair $(\{i,j\},\{k,l\})$ of edges such that $i<k<j<l$. Usually, one extends the triangular array A to a matrix by setting $a_{j,i}=-a_{i,j}$, $i<j$, and $a_{i,i}=0$ for all i. Then, abusing notation, we identify the triangular array with the skew-symmetric matrix $A=(a_{i,j})_{1\le i,j\le 2n}$. The Pfaffian satisfies the following useful properties:

$$\operatorname{Pf}(B^tAB)=\det(B)\operatorname{Pf}(A)$$

and

$$\mathrm{Pf}(A)^2 = \det(A). \tag{10.2}$$

The latter equality shows in particular that Pfaffians are very close to determinants. They do, in fact, generalize determinants since

$$\mathrm{Pf}\begin{pmatrix} 0 & B \\ -B & 0 \end{pmatrix} = \det B, \tag{10.3}$$

for any square matrix B.

Thus, given a graph with vertices $v_1, v_2, \ldots, v_{2n}$, specializing $a_{i,j}$ to the weight of the edge between v_i and v_j if it exists and setting $a_{i,j} = 0$ otherwise in the definition of the Pfaffian, we obtain almost $M_w(G)$, the only difference is that there could be signs in front of the individual terms of the sum, whereas in $M_w(G)$ the sign in front of each term must be $+$. (The object which one obtains by omitting the sign in (10.1) is called *Hafnian*. Unfortunately, in contrast to the Pfaffian, it does not have any nice properties and it is therefore extremely difficult to compute.) Kasteleyn's idea is to circumvent this problem by orienting the edges of the graph, defining signed weights of the edges, in such a way that the Pfaffian of the array with signed weights produces exactly $M_w(G)$.

More precisely, given a (weighted) graph G with vertices $v_1, v_2, \ldots, v_{2n}$, we make it into an oriented (weighted) graph $\overrightarrow{G}$. That is, if there is an edge between v_i and v_j, $e_{i,j}$ say, we orient it either from v_i to v_j or the other way. Now we define the *signed adjacency matrix* $A(\overrightarrow{G})$ of $\overrightarrow{G}$ by letting its (i,j)-entry to be $+w(e_{i,j})$ if there is an edge from v_i to v_j oriented that way, $-w(e_{i,j})$ if there is an edge from v_j to v_i oriented that way, and 0 if there is no edge between v_i and v_j. Such an orientation is called *Pfaffian* if

$$\mathrm{Pf}(A(\overrightarrow{G})) = \pm M_w(G).$$

Clearly, the question remains whether a Pfaffian orientation can be found for a given graph. In general, this is an open question. However, Kasteleyn shows that for *planar* graphs such a Pfaffian orientation can always be found. Moreover, he shows that any orientation of a planar graph which has the property that around any face bounded by $4k$ edges an odd number of edges is oriented in either direction and that around any face bounded by $4k + 2$ edges an even number of edges is oriented in either direction is Pfaffian.

For *bipartite* graphs (that is, for graphs in which the set of vertices can be split into two disjoint sets such that all the edges connect the vertex of one of these sets to a vertex of the other), the situation is even nicer. This is because for a bipartite graph G in which both parts of the bipartition of the vertices are of the same size (otherwise there is no perfect matching), any signed adjacency matrix $A(\overrightarrow{G})$ has the block form of the matrix on the left-hand side of (10.3) and, hence, the Pfaffian reduces to a determinant. More precisely, let G be a bipartite graph with vertex set $V = U \cup W$, $U = \{u_1, u_2, \dots, u_n\}$ and $W = \{w_1, w_2, \dots, w_n\}$, with edges connecting some u_i to some w_j. Given a Pfaffian orientation $\overrightarrow{G}$, we build the *signed bipartite adjacency matrix* $B(\overrightarrow{G}) = (b_{i,j})_{1\le i,j\le n}$ of $\overrightarrow{G}$ by setting $b_{i,j} = +w(e_{i,j})$ if there is an edge from u_i to w_j oriented that way, $-w(e_{i,j})$ if there is an edge from u_j to w_i oriented that way, and 0 if there is no edge between u_i and w_j. Then we have

$$\det(B(\overrightarrow{G})) = \pm M_w(G).$$

In particular, this holds for any bipartite *planar* graph. See [36] for a structural description which (not necessarily planar) bipartite graphs admit a Pfaffian orientation.

Kasteleyn's construction in the planar case has been generalized to graphs on surfaces of any genus g in [9, 15, 42], independently. As predicted by Kasteleyn, the solution is in terms of a linear combination of 4^g Pfaffians.

With the help of his method, Kasteleyn computed the number of dimer coverings of an $m \times n$ rectangle. (A dimer is a 2×1 rectangle. Thus, this is equivalent to counting the number of perfect matchings on the $m \times n$ grid graph. The formula was independently found by Temperley and Fisher.) The result is

$$\prod_{i=1}^{m}\prod_{j=1}^{n}\left(2\cos\frac{\pi i}{m+1} + 2\sqrt{-1}\cos\frac{\pi j}{n+1}\right).$$

For even m and n, the formula can be rewritten as

$$\prod_{i=1}^{m/2}\prod_{j=1}^{n/2}\left(4\cos^2\frac{\pi i}{m+1} + 4\cos^2\frac{\pi j}{n+1}\right).$$

There is a similar rewriting if one of m and n is odd. (If both m and n are odd, there is no dimer covering.)

For further reading and references see [27].

11 Non-intersecting paths

Let $G = (V, E)$ be a directed acyclic graph with vertices V and directed edges E. Furthermore, we are given a function w which assigns a weight $w(x)$ to every vertex or edge x. Let us define the weight $w(P)$ of a walk P in the graph by $\prod_e w(e) \prod_v w(v)$, where the first product is over all edges e of the walk P and the second product is over all vertices v of P. We denote the set of all walks in G from u to v by $\mathcal{P}(u \to v)$, and the set of all families $(P_1, P_2, \dots, P_n)$ of walks, where P_i runs from u_i to v_i, $i = 1, 2, \dots, n$, by $\mathcal{P}(\mathbf{u} \to \mathbf{v})$, with $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$. The symbol $\mathcal{P}^+(\mathbf{u} \to \mathbf{v})$ stands for the set of all families $(P_1, P_2, \dots, P_n)$ in $\mathcal{P}(\mathbf{u} \to \mathbf{v})$ with the additional property that no two walks share a vertex. We call such families of walk(er)s *"vicious walkers"* or, alternatively, *"non-intersecting paths"*. The weight $w(\mathbf{P})$ of a family $\mathbf{P} = (P_1, P_2, \dots, P_n)$ of walks is defined as the product $\prod_{i=1}^n w(P_i)$ of all the weights of the walks in the family. Finally, given a set M with weight function w, we write $\mathrm{GF}(\mathcal{M}; w)$ for the generating function $\sum_{x \in \mathcal{M}} w(x)$.

We need two further notations before we are able to state the *Lindström–Gessel–Viennot theorem*. (For referennces and historical remarks, we refer the reader to Footnote 5 in [24].) As earlier, the symbol $\mathfrak{S}_n$ denotes the symmetric group of order n. Given a permutation $\sigma \in \mathfrak{S}_n$, we write $\mathbf{u}_\sigma$ for $(u_{\sigma(1)}, u_{\sigma(2)}, \dots, u_{\sigma(n)})$. Then

$$\sum_{\sigma \in \mathfrak{S}_n} (\operatorname{sgn} \sigma) \cdot \mathrm{GF}(\mathcal{P}^+(\mathbf{u}_\sigma \to \mathbf{v}); w) = \det_{1 \le i,j \le n} \big(\mathrm{GF}(\mathcal{P}(u_j \to v_i); w) \big). \tag{11.1}$$

Most often, this theorem is applied in the case where the only permutation σ for which vicious walks exist is the identity permutation, so that the sum on the left-hand side reduces to a single term which counts all families $(P_1, P_2, \dots, P_n)$ of vicious walks, the i-th walk P_i running from A_i to E_i, $i = 1, 2, \dots, n$. This case occurs for example if for any pair of walks (P, Q) with P running from u_a to v_d and Q running from u_b to v_c, $a < b$ and $c < d$, it is true that P and Q must have a common vertex. Explicitly, in that case we have

$$\mathrm{GF}(\mathcal{P}^+(\mathbf{u} \to \mathbf{v}); w) = \det_{1 \le i,j \le n} \big(\mathrm{GF}(\mathcal{P}(u_j \to v_i); w) \big). \tag{11.2}$$

If the starting points or/and the end points are not fixed, then the corresponding number is given by a Pfaffian, a result obtained by Okada and Stembridge (see [6] for references). For a set $\mathcal{A}$ of starting points, let $\mathcal{P}^+(\mathcal{A} \to \mathbf{v})$

denote the set of all families $(P_1, P_2, \dots, P_{2n})$ of non-intersecting lattice paths, where P_i runs from some point of $\mathcal{A}$ to v_i, $i = 1, 2, \dots, 2n$. Furthermore, let us suppose that the elements of $\mathcal{A} = \{u_1, u_2, \dots\}$ are ordered in such a way that for any pair of walks (P, Q) with P running from u_a to v_d and Q running from u_b to v_c, $a < b$ and $c < d$, it is true that P and Q must have a common vertex. (This is the same condition as the one which makes (11.2) valid, with the only difference that, here, the number of u_i's could be larger than the number of v_i's.) Then,

$$\begin{aligned} &\mathrm{GF}(\mathcal{P}^+(\mathcal{A} \to \mathbf{v}); w) \\ &\qquad = \operatorname*{Pf}_{1 \le i,j \le 2n} \Bigg(\sum_{a<b} \big(\mathrm{GF}(\mathcal{P}(u_a \to v_i); w)\, \mathrm{GF}(\mathcal{P}(u_b \to v_j); w) \\ &\qquad\qquad - \mathrm{GF}(\mathcal{P}(u_b \to v_i); w)\, \mathrm{GF}(\mathcal{P}(u_a \to v_j); w) \big) \Bigg). \qquad (11.3) \end{aligned}$$

If the number of paths is odd, then one can use the same formula by adding an artificial point to the end points and to the set of starting points $\mathcal{A}$. There is also a theorem by Okada and Stembridge which covers the case that starting *and* end points vary. Refinements when the number of turns is fixed can be found in [22].

12 Vicious walkers, plane partitions, rhombus tilings, and fully packed loop configurations

In this section we describe the interrelations between four frequently appearing objects in statistical mechanics and combinatorics: vicious walkers, plane partitions, rhombus tilings, and fully packed loop configurations.

Given a lattice, *vicious walkers*, as introduced by Fisher [11], are particles which move on lattice sites in such a way that two particles never occupy the same lattice site. Models of vicious walkers have been the object of numerous studies from various points of view. Rather than accomplishing the impossible task of providing a complete overview of references, the reader is referred to the basic reference [11] and to [24] for further pointers to the literature.

Most of the known results apply for vicious walkers on the line. There are in fact two different models: in the *random turns vicious walker model*,

n walkers move on the integral points of the real line in such a way that at each tick of the clock *exactly one* walker moves to the right or to the left, whereas in the *lock step vicious walker model* n walkers move on the integral points of the real line in such a way that at each tick of the clock *each* walker moves to the right or to the left.

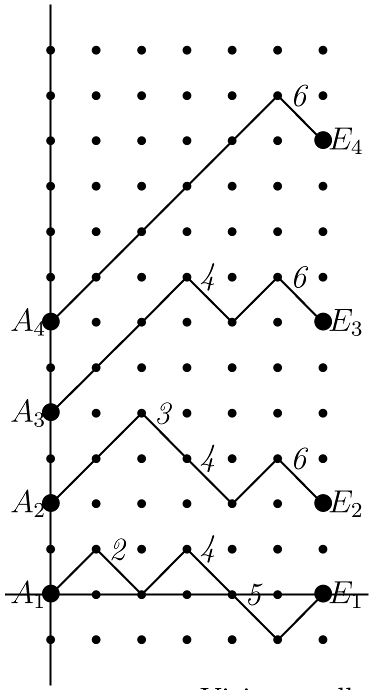
The first model is equivalent to a model of one walker in $\mathbb{Z}^n$ ($\mathbb{Z}$ denoting the set of integers) which at each tick of the clock moves a positive or negative unit step in direction of one of the coordinate axes, always staying in the wedge $x_1 > x_2 > \cdots > x_n$. This point of view was already put forward in [11]. However, this problem belongs to the problem of counting paths in chambers of reflection groups discussed in Section 9.

The second model could also be realized as a single walker model (cf. [23]). However, most often it is realized as a model of n paths in the plane consisting of steps $(1,1)$ and $(1,-1)$ with the property that no two paths have a point in common. In this picture, the x-axis becomes the time line, the k-th path doing an up-step $(1,1)$ from $(t-1,y)$ to $(t,y+1)$ meaning that the k-th particle moves to the left at time t, whereas the k-th path doing a down-step $(1,-1)$ from $(t-1,y)$ to $(t,y-1)$ meaning that the k-th particle moves to the right at time t.

The reader should consult Figure 14.a for an example. (The labellings should be ignored at this point.) Clearly, what we encounter here is a particular instance of the non-intersecting paths of Section 11. Therefore, for fixed starting and end points, Formula (11.2) applies, whereas if the starting points vary and the end points are fixed, it is Formula (11.3) which applies.

At this point, the links to the other objects, semistandard tableaux and plane partitions (cf. [6]), emerge. A filling of the cells of the Ferrers diagram of λ with elements of the set $\{1,2,\dots\}$ which is *weakly increasing* along rows and *strictly increasing* along columns is called a (*semistandard*) *tableau of shape* λ. Figure 14.b shows such a semistandard tableau of shape $(4,3,2)$. In fact, vicious walkers and semistandard tableaux are equivalent objects. To see this, first label down-steps by the x-coordinate of their end point, so that a step from $(a-1,b)$ to $(a,b-1)$ is labelled by a, see Figure 14.a. Then, out of the labels of the j-th path, form the j-th column of the corresponding tableau, see Figure 14.b. The resulting array of numbers is indeed a semistandard tableau. This can be readily seen, since the entries are trivially strictly increasing along columns, and they are weakly increasing along rows because the paths do not touch each other. Thus, problems of enumerating vicious walkers can be translated into tableau enumeration problems, and vice versa.

The significance of semistandard tableaux lies particularly in the repre-



a. Vicious walkers

2	3	4	6
4	4	6	
5	6		

b. A tableau

Figure 14:

sentation theory for classical groups. Namely, the irreducible characters for $GL(n,\mathbb{C})$ and $SL(n,\mathbb{C})$, the *Schur functions*, are generating functions for semistandard tableaux of a given shape. If the entries of the i-th row of a semistandard tableau are required to be at least $2i-1$, then one speaks of *symplectic tableaux*, and the irreducible characters for $Sp(2n,\mathbb{C})$ are generating functions for symplectic tableaux of a given shape. We refer the reader to [26] for more information on these topics.

Objects which are very close to semistandard tableaux are *plane partitions*. According to MacMahon, a plane partition of shape λ is a filling of the Ferrers diagram of λ with non-negative integers which is *weakly decreasing* along rows and columns. See Figure 15.b for an example of a plane partition of shape $(3,3,3)$. In particular, semistandard tableaux and plane partitions of rectangular shape are actually equivalent. For, let T be a semistandard tableau of rectangular shape. Then, from each element of the i-th row we subtract i. Finally, the obtained array is rotated by $180°$. As a result, we obtain a plane partition. See Figure 15 for a semistandard tableau and a plane partition which correspond to each other under these transformations.

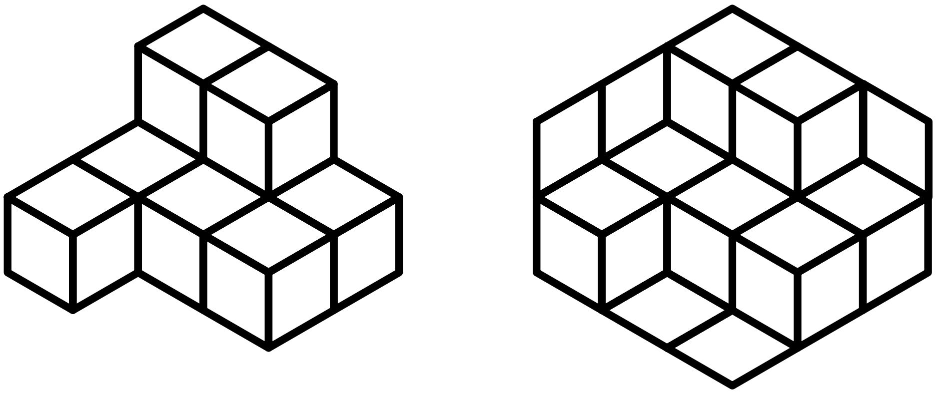
1	1	2
3	3	3
4	5	5

a. A semistandard tableau

2	2	1
1	1	1
1	0	0

b. A plane partition

Figure 15:



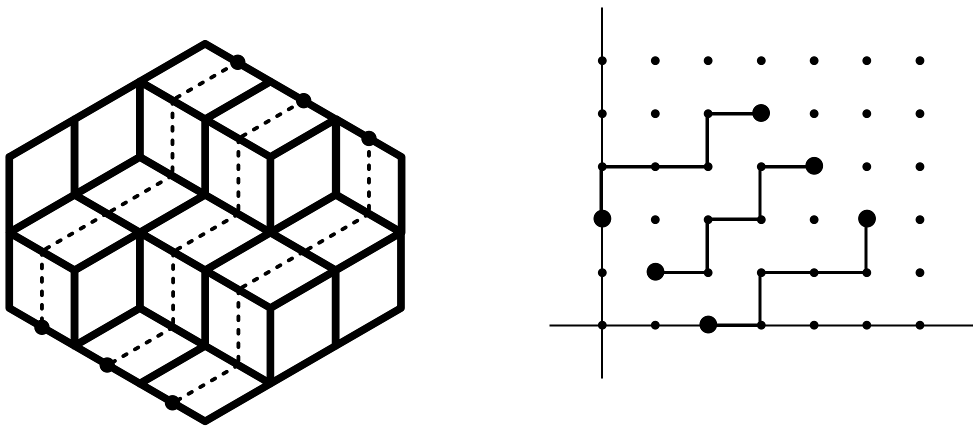
a. A plane partition, 3-dimensional b. A rhombus tiling

Figure 16:

On the other hand, plane partitions can also be realized as 3-dimensional objects, by interpreting each entry in the array as a pile uf unit cubes of the size of the entry. For example, the plane partition in Figure 15 corresponds to the pile of cubes in Figure 16.a. But then, forgetting the 3-dimensional view, by embedding the picture in a minimally bounding hexagon, and by filling the emerging empty regions by rhombi of unit length in the unique way this is possible, we obtain a *rhombus tiling* of a hexagon in which opposite sides have the same length, see Figure 16.b.

From the rhombus tiling, there is then again an elegant way to go to nointersecting paths: we mark the mid-points of the edges along two opposite sides, see Figure 17.a. Now we draw lattice paths which connect points on different sides, by "following" along the other lozenges, as is indicated in Figure 17.a by the dashed lines. Clearly, the resulting paths are nonintersecting, i.e., no two paths have a common vertex. If we slightly distort the underlying lattice, we get orthogonal paths with horizontal and vertical steps in the positive direction, see Figure 17.b.

Rhombus tilings, on their part, are equivalent to perfect matchings of hexagonal graphs. To see this, one places the tiling on the underlying trian-



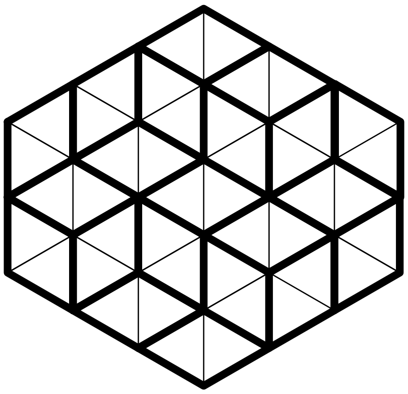
a. A rhombus tiling b. A family of non-intersecting paths

Figure 17:

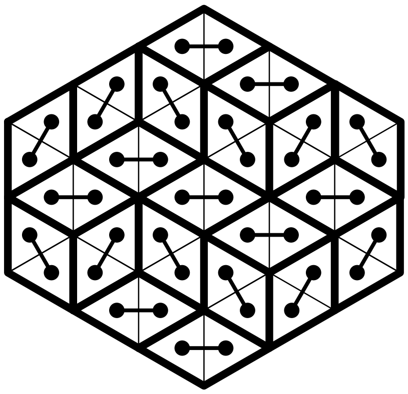
gular grid, see Figure 18.a. Then one places a bond into each rhombus, so that it connects the mid-points of the two triangles out of which the rhombus is composed, see Figure 18.b. Finally, one forgets the contour of the tiling, but instead one introduces all the other edges which connect mid-points of adjacent triangles of the triangular grid, see Figure 18.c. Thus, one arrives at a perfect matching of the hexagonal graph consisting of the edges connecting mid-points of triangles.

Because of these various connections, enumeration problems for vicious walkers, plane partitions, tableaux, rhombus tilings can be approached by the different methods which are available for the various objects: the determinant theorem from Section 11, together with determinant evaluation techniques (cf. the survey [25]), apply, as well as the "Kasteleyn method" from Section 10, and also methods from character theory for the classical groups. And all of them have been applied extensively (see the surveys [20, 34] and [26]), the first and third more frequently for exact enumeration, while the second particularly for asymptotic studies. It should be noted that methods from random matrix theory also apply in certain situations, see [19].

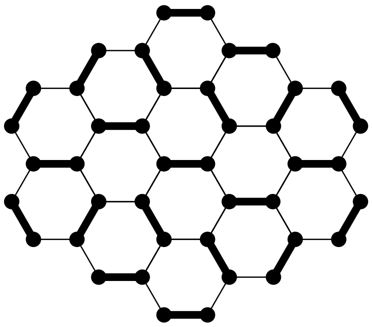
We missed in fact to mention a further object, from statistical physics, which in some cases is equivalent to vicious walkers, etc.: *fully packed loop configurations*. (Fully packed loop configurations are in bijection with six vertex configurations, see Section 13.) If one imposes certain "connectivity constraints" on fully packed loop configurations, then one can construct bijections with rhombus tilings and, hence, with nonintersecting paths and with the other objects discussed in this section. The reader is referred to [13]



a. A rhombus tiling



b. Bonds in rhombi



c. A perfect matching of a hexagonal graph

Figure 18:

and the references therein.

After having explained the various connections, we cite some fundamental results in the area. (We refer the reader to [6] and [40, Ch. 7].) MacMahon proved that the number of all plane partitions contained in an $a \times b \times c$ box (when viewed in 3 dimensions) is equal to

$$\prod_{i=1}^{a}\prod_{j=1}^{b}\prod_{k=1}^{c}\frac{i+j+k-1}{i+j+k-2}. \tag{12.1}$$

Thus, the number of rhombus tilings of a hexagon with side lengths a, b, c, a, b, c is given by the same number, as well as the number of all vicious walkers $(P_1, P_2, \dots, P_a)$, where P_i runs from $(0, 2i)$ to $(b+c, b-c+2i)$, $i = 1, 2, \dots, a$. More generally, the number of semistandard tableaux of shape λ with entries at most m is given by the *hook-content formula*

$$\prod_{u\in\lambda}\frac{c(u)+m}{h(u)}, \tag{12.2}$$

where u ranges over all the cells of the Ferrers diagram of λ, with $c(u)$ being the *content* of u, defined as the difference of the column number and the row number of u, and with $h(u)$ being the *hook length* of u, defined as the number of cells in the *hook* of u, the latter consisting of the cells to the right of u in the same row and below u in the same column, including u. Thus, this gives also a formula for the number of all vicious walkers $(P_1, P_2, \dots, P_a)$, where P_i runs from $(0, 2i)$ to (N, h_i). See [26, Sec. 2] for details. There it is also explained that a Schur function summation formula, together with an analogue of the hook-content formula for *special orthogonal characters*, proves that the number of all the number of all vicious walkers $(P_1, P_2, \dots, P_a)$, where P_i runs from $(0, 2i)$ for N steps is given by

$$\prod_{1\le i\le j\le N}\frac{a+i+j-1}{i+j-1}. \tag{12.3}$$

The reader is referred to the references given in this section for many more results, in particular, on the enumeration of plane partitions with symmetry, the enumeration of rhombus tilings of regions other than hexagons, and the enumeration of vicious walkers with various starting and end points, under various constraints.

13 Six vertex model and alternating sign matrices

An alternating sign matrix is a square matrix of 0's, 1's and -1's for which the sum of entries in each row and in each column is 1 and the non-zero entries of each row and of each column alternate in sign. For instance,

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

is a 5×5 alternating sign matrix. Zeilberger proved that the number of $n \times n$ alternating sign matrices is given by

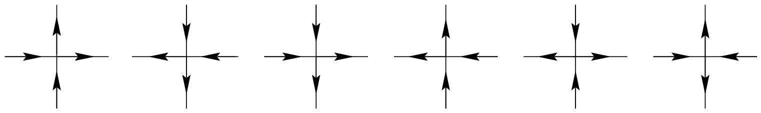
$$\prod_{i=0}^{n-1} \frac{(3i+1)!}{(n+i)!}, \tag{13.1}$$

and he went on to prove the finer version that the number of $n \times n$ alternating sign matrices with the (unique) 1 in the first row in position j is given by

$$\frac{\binom{n+j-2}{n-1}\binom{2n-j-1}{n-1}}{\binom{3n-2}{n-1}} \prod_{i=0}^{n-1} \frac{(3i+1)!}{(n+i)!}. \tag{13.2}$$

The first number is also equal to the number of totally symmetric self-complementary plane partitions contained in the $(2n) \times (2n) \times (2n)$ box, but there is no intrinsic explanation why this is so. We refer the reader to [6] for an exposition of these results, and for pointers to the literature containing further unexplained connections between alternating sign matrices and plane partitions.

While the first result was achieved by a brute force constant term approach, the second result is based on the observation, that alternating sign matrices are in bijection with configurations in the six vertex model on the square grid under domain wall boundary conditions. This then allowed one to use a formula due to Izergin for the partition function for these six vertex configurations. Similar formulas for variations of the model have been found by Kuperberg, and by Razumov and Stroganov (see [35] and the references therein).

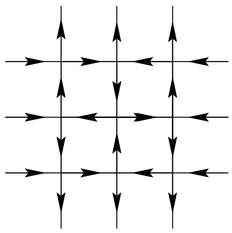


0 0 0 0 -1 1

Figure 19:

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

a. An alternating sign matrix



b. A six vertex configuration

Figure 20:

A configuration in the six vertex model is an orientation of edges of a 4-regular graph (that is, at each vertex there meet exactly 4 edges) such that at each vertex two edges are oriented towards the vertex and two edges are oriented away from the vertex. Thus, there are six possible vertex configurations, giving the name of the model, see Figure 19. To go from one object to the other one uses the translation between local configurations at a vertex and entries in alternating sign matrices indicated in the figure. An example of the correspondence can be found in Figure 20.

Another manifestation of alternating sign matrices and six vertex configurations are *fully packed loop configurations.* A fully packed loop configuration on a graph is a collection of edges such that each vertex is incident to exactly two edges. One obtains a fully packed loop configuration out of a six vertex configuration by dividing the square lattice into its even and odd sublattice denoted by A and B respectively. Instead of arrows, only those edges are drawn that on sublattice A point inward and on sublattice B point outward. I refer to [17, 13] for further reading.

The story of alternating sign matrices and its connection to the six vertex model is given a vivid account in [6], with further important results by Kuperberg, Okada, Razumov and Stroganov, referenced in [35].

Fully packed loop configurations seem to play an important role in the explicit form of the groundstate vectors of certain Hamiltonians in the dense

$O(1)$ loop model. The corresponding conjectures are surveyed in [17]. There is important progress on these conjectures by Di Francesco and Zinn–Justin (cf. [14] and the references cited therein).

14 Binomial sums and hypergeometric series

When dealing with enumerative problems, it is inevitable to deal with *binomial sums*, that is, sums in which the summands are products/quotients of binomial coefficients and factorials, such as, for example,

$$\sum_{k=0}^{n} \binom{2k}{k} \binom{2n-2k}{n-k}.$$

In most cases, the right environment in which one should work is the theory of (generalized) *hypergeometric series*. These are defined as follows:

$${}_rF_s\left[\begin{matrix} a_1,\dots,a_r \\ b_1,\dots,b_s \end{matrix}; z\right] = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_r)_k}{(b_1)_k \cdots (b_s)_k} \frac{z^k}{k!} ,$$

where $(\alpha)_k = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+k-1)$ for $k > 0$, and $(\alpha)_0 = 1$. The symbol $(\alpha)_k$ is called the *Pochhammer symbol* or *shifted factorial.* For in-depth treatments of the subject, we refer the reader to [2, 16, 38].

Hypergeometric series can be characterized as series in which the quotient of the $(k+1)$-st by the k-th summand is a rational function in k. This is also the way to convert binomial sums into their hypergeometric form (respectively to see if this is possible; in most cases it is): form the quotient of the $(k+1)$-st by the k-th summand and read off the parameters $a_1,\dots,a_r$, $b_1,\dots,b_s$, and the argument z from the factorization of the numerator and the denominator polynomials of the rational function, out of these form the corresponding hypergeometric series, and multiply the series by the summand for $k = 0$. This is, in fact, a completely routine task, and, indeed, computer algebra programs such as *Maple* and *Mathematica* do this automatically.

The reason why hypergeometric series are much more fundamental than the binomial sums themselves is that there are hundreds of ways to write the same sum using binomial coefficients and factorials, whereas there is *just one* hypergeometric form, that is, hypergeometric series are a kind of *normal form* for binomial sums. In particular, given a specific binomial sum, it is a hopeless enterprise to scan through all the identities available in the literature for this sum. There may be an identity for it, but perhaps written in a different way. On the contrary, given a specific hypergeometric series, the list of available identities which apply to this series is usually not large, and tables of such identities can be set up in a systematic way. This has been done (cf. [38]; the most comprehensive table available to this date is contained in the manual of the *Mathematica* package HYP[5]), and scanning through these tables is largely facilitated by the use of the *Mathematica* package HYP[5].

We give here some of the most important identities for hypergeometric series. Aside from the binomial theorem, the most important *summation formulas* are *Gauß' ${}_2F_1$-summation formula*

$$ {}_2F_1\left[\begin{matrix}a,b\\ c\end{matrix};1\right]=\frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}, $$

provided $\Re(c-a-b)>0$, the *Pfaff–Saalschütz summation formula*

$$ {}_3F_2\left[\begin{matrix}a,b,-n\\ c,1+a+b-c-n\end{matrix};1\right]=\frac{(c-a)_n\,(c-b)_n}{(c)_n\,(c-a-b)_n}, $$

provided n is a non-negative integer, and *Dougall's summation formula*

$$ {}_7F_6\left[\begin{matrix}a,a/2+1,b,c,d,1+2a-b-c-d+n,-n\\ a/2,1+a-b,1+a-c,1+a-d,-a+b+c+d-n,a+1+n\end{matrix};1\right] \\ =\frac{(1+a)_n\,(1+a-b-c)_n\,(1+a-b-d)_n\,(1+a-c-d)_n}{(1+a-b)_n\,(1+a-c)_n\,(1+a-d)_n\,(1+a-b-c-d)_n}, $$

provided n is a non-negative integer.

Some of the most important *transformation formulas* are *Euler's transformation formula*

$$ {}_2F_1\left[\begin{matrix}a,b\\ c\end{matrix};z\right]=(1-z)^{c-a-b}{}_2F_1\left[\begin{matrix}c-a,c-b\\ c\end{matrix};z\right], $$

provided $|z|<1$, *Kummer's transformation formula*

$$ {}_3F_2\left[\begin{matrix}a,b,c\\ d,e\end{matrix};1\right]=\frac{\Gamma(e)\,\Gamma(d+e-a-b-c)}{\Gamma(e-a)\,\Gamma(d+e-b-c)}{}_3F_2\left[\begin{matrix}a,d-b,d-c\\ d,d+e-b-c\end{matrix};1\right], $$

provided both series converge, and *Whipple's transformation formula*

$$ {}_4F_3\left[\begin{matrix}a,b,c,-n\\ e,f,1+a+b+c-e-f-n\end{matrix};1\right]=\frac{(e-a)_n\,(f-a)_n}{(e)_n\,(f)_n} \\ \times{}_4F_3\left[\begin{matrix}-n,a,1+a+c-e-f-n,1+a+b-e-f-n\\ 1+a+b+c-e-f-n,1+a-e-n,1+a-f-n\end{matrix};1\right] \qquad (14.1) $$

[5]available at `http://www.mat.univie.ac.at/~kratt`

where n is a non-negative integer, and

$$
{}_7F_6\left[\begin{matrix} a, 1+\frac{a}{2}, b, c, d, e, -n \\ \frac{a}{2}, 1+a-b, 1+a-c, 1+a-d, 1+a-e, 1+a+n \end{matrix}; 1\right]
$$
$$
= \frac{(1+a)_n\,(1+a-d-e)_n}{(1+a-d)_n\,(1+a-e)_n}\,{}_4F_3\left[\begin{matrix} 1+a-b-c, d, e, -n \\ 1+a-b, 1+a-c, -a+d+e-n \end{matrix}; 1\right], \tag{14.2}
$$

provided n is a non-negative integer.

Since about 1990, for the *verification* of binomial and hypergeometric series, there are *automatic tools* available. The book [33] is an excellent introduction into these aspects. The philosophy is as follows. Suppose we are given a binomial or hypergeometric series $S(n) = \sum_k F(n,k)$. The *Gosper–Zeilberger algorithm*[6] (cf. [33]; a simplified version was presented in [45]) will find a linear recurrence

$$
A_0(n)S(n) + A_1(n)S(n+1) + \cdots + A_d(n)S(n+d) = C(n) \tag{14.3}
$$

for some d, where the coefficients $A_i(n)$ are polynomials in n, and where $C(n)$ is a certain function in n, *with proof*!

If, for example, we suspected that $S(n) = RHS(n)$, where $RHS(n)$ is some closed form expression, then we just have to verify that $RHS(n)$ satisfies the recurrence (14.3) and check $S(n) = RHS(n)$ for sufficiently many initial values of n to have a proof for the identity $S(n) = RHS(n)$ *for all* n. On the other hand, if $RHS(n)$ was a different sum, then we would apply the algorithm to find a recurrence for $RHS(n)$. If it turns out to be the same recurrence then, again, a check of $S(n) = RHS(n)$ for a few initial values will provide a full proof of $S(n) = RHS(n)$ *for all* n.

Even in the case that we do not have a conjectured expression $RHS(n)$, this is not the end of the story. Given a recurrence of the type (14.3), the *Petkovšek algorithm*[7] (cf. [33]) is able to find a closed form solution (where "closed form" has a precise meaning, see [33]), respectively tell that there is no closed form solution.

The fascinating point about both algorithms is that neither do we have to know what the algorithm does internally nor do we have to check that.

[6] Computer implementations written by Doron Zeilberger are available from `http://www.math.rutgers.edu/~zeilberg`, *Mathematica* implementations written by Peter Paule and Markus Schorn are available from `http://www.risc.uni-linz.ac.at/research/combinat/risc/software`.

[7] A *Mathematica* implementation by Marko Petkovšek is available at `http://www.fmf.uni-lj.si/~petkovsek`.

For the Petkovšek algorithm this is obvious anyway because, once the computer says that a certain expression is a solution of (14.3), it is a routine matter to check that. This is less obvious for the Gosper–Zeilberger algorithm. However, what the Gosper–Zeilberger algorithm does is, for a given sum $S(n) = \sum_k F(n,k)$, to find polynomials $A_0(n), A_1(n), \dots, A_d(n)$ *and* an expression $G(n,k)$ (which is in fact a rational multiple of $F(n,k)$), such that

$$A_0(n)F(n,k)+A_1(n)F(n+1,k)+\dots+A_d(n)F(n+d,k) = G(n,k+1)-G(n,k), \tag{14.4}$$

for some d. Because of the properties of $F(n,k)$ and $G(n,k)$, which are part of the theory, this is an identity which can be *directly* verified by clearing all common factors and checking the remaining identity between rational functions in n and k. However, we may now sum both sides of (14.4) over k to obtain a recurrence of the form (14.3).

Algorithms for multiple sums are also available[8]. They follow ideas by Wilf and Zeilberger [44] (of which a simplified version is presented in [29]), however they run more quickly in capacity problems. Schneider [37] is currently developing a very promising new algorithmic approach to the automatic treatment of multisums.

References

[1] G. E. Andrews, *The theory of partitions*, Encyclopedia of Math. and its Applications, Vol. 2, Addison–Wesley, Reading, 1976; reprinted by Cambridge University Press, Cambridge, 1998.

[2] G. E. Andrews, R. A. Askey et R. Roy, *Special Functions*, The Encyclopedia of Mathematics and Its Applications, vol. 71, G.-C. Rota, ed., Cambridge University Press, Cambridge, 1999.

[3] R. Ayoub, *An introduction to the analytic theory of numbers*, Mathematical Surveys, No. 10, Amer. Math. Soc., Providence, R.I., 1963.

[4] F. Bergeron, G. Labelle and P. Leroux, *Combinatorial species and tree-like structures*, Cambridge University Press, Cambridge, 1998.

[5] M. Bousquet-Mélou and A. Jehanne, *Polynomial equations with one catalytic variable, algebraic series, and map enumeration*, preprint; `ar`χ`iv:math.CO/0504018`.

[6] D. M. Bressoud, *Proofs and confirmations — The story of the alternating sign matrix conjecture*, Cambridge University Press, Cambridge, 1999.

[8] *Mathematica* implementations by Axel Riese and Kurt Wegschaider are available at `http://www.risc.uni-linz.ac.at/research/combinat/risc/software`.

[7] N. G. de Bruijn, *Pólya's theory of counting*, in: Applied Combinatorial Mathematics, E. .F. Beckenbach, ed., Wiley, New York, 1964; reprinted by Krieger, Malabar, Florida, 1981.

[8] L. Comtet, *Advanced Combinatorics*, D. Reidel, Dordrecht, Holland, 1974.

[9] N. P. Dolbilin, A. S. Mishchenko, M. A. Shtan'ko, M. I. Shtogrin and Yu. M. Zinoviev, *Homological properties of dimer configurations for lattices on surfaces* Funct. Anal. Appl. **30** (1996), 163–173.

[10] W. Feller, *An introduction to probability theory and its applications*, vol. 1, 2nd ed., Wiley, New York, 1957.

[11] M. E. Fisher, *Walks, walls, wetting and melting*, J. Stat. Phys. **34** (1984), 667-729.

[12] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, book project, available at `http://algo.inria.fr/flajolet/Publications/books.html`.

[13] P. Di Francesco, P. Zinn-Justin and J.-B. Zuber, *Determinant Formulae for some Tiling Problems and Application to Fully Packed Loops*, preprint; arχiv:math-ph/0410002.

[14] P. Di Francesco and P. Zinn-Justin, *Quantum Knizhnik-Zamolodchikov equation, generalized Razumov-Stroganov sum rules and extended Joseph polynomials*, preprint; arχiv:math-ph/0508059.

[15] A. Galluccio and M. Loebl, *On the theory of Pfaffian orientations I. Perfect matchings and permanents*, Electron. J. Combin. **6** (1999), Article #R6, 18 pp.

[16] G. Gasper and M. Rahman, *Basic hypergeometric series*, 2nd edition, Encyclopedia of Mathematics And Its Applications 96, Cambridge University Press, Cambridge, 2004.

[17] J. de Gier, *Loops, matchings and alternating-sign matrices*, Discr. Math., to appear. arχiv:math.CO/0211285.

[18] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge University Press, Cambridge, 1990.

[19] K. Johansson, *Non-intersecting paths, random tilings and random matrices*, Probab. Theory Related Fields **123** (2002), 225–280.

[20] R. Kenyon, *An introduction to the dimer model*, lecture notes for a short course at the ICTP, 2002; arχiv:math.CO/0310326.

[21] R. Koekoek and R. F. Swarttouw, *The Askey–scheme of hypergeometric orthogonal polynomials and its q-analogue*, TU Delft, The Netherlands, 1998; on the WWW: `http://aw.twi.tudelft.nl/~koekoek/askey.html`.

[22] C. Krattenthaler, *The enumeration of lattice paths with respect to their number of turns*, in: Advances in Combinatorial Methods and Applications to Probability and Statistics, N. Balakrishnan, ed., Birkhäuser, Boston, 1997, pp. 29–58.

[23] C. Krattenthaler, *Asymptotics for random walks in alcoves of affine Weyl groups*, preprint. arχiv:math.CO/0301203.

[24] C. Krattenthaler, *Watermelon configurations with wall interaction: exact and asymptotic results*, preprint. `ar`χ`iv:math.CO/0506323`.

[25] C. Krattenthaler, *Advanced determinant calculus: a complement*, Linear Algebra Appl. (to appear); `ar`χ`iv:math.CO/0503507`.

[26] C. Krattenthaler, A. J. Guttmann and X. G. Viennot, *Vicious walkers, friendly walkers and Young tableaux II: with a wall*, J. Phys. A: Math. Gen. **33** (2000), 8835–8866.

[27] G. Kuperberg, *An exploration of the permanent-determinant method*, Electron. J. Combin. **5** (1998), Article #R46, 34 pp.

[28] G. Labelle and C. Lamathe, *A shifted asymmetry index series*, Adv. Appl. Math. **32** (2004), 576–608.

[29] M. Mohammed and D. Zeilberger, *Multi-variable Zeilberger and Almkvist–Zeilberger algorithms and the sharpening of Wilf-Zeilberger theory*, preprint; available at `http://www.math.rutgers.edu/~zeilberg`.

[30] S. G. Mohanty, *Lattice Path Counting and Applications*, Academic Press, New York, 1979.

[31] A. M. Odlyzko, Asymptotic enumeration methods, in: "Handbook of Combinatorics," vol. 2, R. L. Graham, M. Grötschel, L. Lovász, eds., Elsevier, Amsterdam, 1995, pp. 1063–1229.

[32] R. Pemantle and M. C. Wilson, *Twenty combinatorial examples of asymptotics derived from multivariate generating functions*, preprint; available at `http://www.cs.auckland.ac.nz/~mcw`.

[33] M. Petkovšek, H. Wilf and D. Zeilberger, *A=B*, A. K. Peters, Wellesley, 1996.

[34] J. Propp, *Enumeration of matchings: Problems and progress*, in: "New Perspectives in Algebraic Combinatorics", L. Billera, A. Björner, C. Greene, R. Simion, and R. P. Stanley, eds., Mathematical Sciences Research Institute Publications, vol. 38, Cambridge University Press, 1999, pp. 255–291.

[35] A. V. Razumov and Y. G. Stroganov, *Enumeration of quarter-turn symmetric alternating-sign matrices of odd order*, preprint; arχiv:math-ph/0507003.

[36] N. Robertson, P. D. Seymour and R. Thomas, *Permanents, Pfaffian orientations, and even directed circuits*, Ann. Math. (2) **150** (1999), 929–975.

[37] C. Schneider, *A new Sigma approach to multi-summation*, Adv. Appl. Math. **34**(4) (2005), 740–767.

[38] L. J. Slater, *Generalized Hypergeometric Functions*, Cambridge University Press, Cambridge, 1966.

[39] R. P. Stanley, *Enumerative Combinatorics*, Vol. 1, Wadsworth & Brooks/Cole, Pacific Grove, California, 1986; reprinted by Cambridge University Press, Cambridge, 1998.

[40] R. P. Stanley, *Enumerative Combinatorics*, vol. 2, Cambridge University Press, Cambridge, 1999.

[41] G. Szegő, *Orthogonal polynomials*, American Mathematical Society Colloquium Publications Vol. XXIII, New York, 1959.

[42] G. Tesler, *Matchings in graphs on non-oriented surfaces*, J. Combin. Theory Ser. B **78** (2000), 198–231.

[43] X. Viennot and W. James, *Heaps of segments, q-Bessel functions in square lattice enumeration and applications in quantum gravity*, preprint.

[44] H. S. Wilf and D. Zeilberger, *An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities*, Invent. Math. **108** (1992), 575–633.

[45] D. Zeilberger, *Deconstructing the Zeilberger algorithm*, J. Difference Equations Appl. (to appear).