A Combinatorial Miscellany

by

Anders Björner
Matematiska Institutionen
Kungl. Tekniska Högskolan
S-100 44 Stockholm
SWEDEN

Richard P. Stanley *)
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
U.S.A.

September 5, 2010

*) Partially supported by NSF grants #DMS-9500714 and #DMS-9988459.

Contents

Introduction

A recent newcomer to the center stage of modern mathematics is the area called *combinatorics*. Although combinatorial mathematics has been pursued since time immemorial, and at a reasonable scientific level at least since Leonhard Euler (1707–1783), the subject has come into its own only in the last few decades. The reasons for the spectacular growth of combinatorics come both from within mathematics itself and from the outside.

Beginning with the outside influences, it can be said that the recent development of combinatorics is somewhat of a cinderella story. It used to be looked down on by "mainstream" mathematicians as being somehow less respectable than other areas, in spite of many services rendered to both pure and applied mathematics. Then along came the prince of computer science with its many mathematical problems and needs — and it was combinatorics that best fitted the glass slipper held out.

The developments within mathematics that have contributed to the current strong standing of combinatorics are more difficult to pinpoint. One is that, after an era where the fashion in mathematics was to seek generality and abstraction, there is now much appreciation of and emphasis on the concrete and "hard" problems. Another is that it has been gradually more and more realized that combinatorics has all sorts of deep connections with the mainstream areas of mathematics, such as (to name the most important ones) algebra, geometry,

probability and topology.

Our aim with this monograph is to give the reader some answers to the questions "What is combinatorics, and what is it good for?" We will do that not by attempting any kind of general survey, but by describing a few selected problems and results in some detail. We want to bring you some examples of problems from "pure" combinatorics, some examples illustrating its interactions with other parts of mathematics, and a few glimpses of its use for computer science. Fortunately, the problems and results of combinatorics are usually quite easy to state and explain, even to the layman with a solid knowledge of high school mathematics. Its accessibility is one of its many appealing aspects. For instance, most popular mathematical puzzles and games, such as Rubik's cube and jigsaw puzzles, are essentially problems in combinatorics.

To achieve our stated purpose it has been necessary to concentrate on a few topics, leaving many of the specialities within combinatorics without mention. The choice will naturally reflect our own interests. The discussion in the Notes section points to some more general accounts that can help remedy this shortcoming.

With some simplification, combinatorics can be said to be the mathematics of the finite. One of the most basic properties of a finite collection of objects is its number of elements. For instance, take words formed from the letters a, b, and c, using each letter exactly once. There are six such words:

$$abc, \quad acb, \quad bac, \quad bca, \quad cab, \quad cba.$$

Now, say that we have n distinct letters. How many words can be formed? The answer is $n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$, because the first letter can be chosen in n ways, then the second one in $n-1$ ways (since the letter already chosen as the first letter is no longer available), the third one in $n-2$ ways, and so on. Furthermore, the total number must be the product of the number of individual choices.

The number of words that can be formed with n letters is an example of an *enumerative* problem. Enumeration is one of the most basic and important aspects of combinatorics. In many branches of mathematics and its applications you need to know the number of ways of doing something.

One of the classical problems of enumerative combinatorics is to count *partitions* of various kinds, meaning the number of ways to break an object into smaller objects of the same kind. Various kinds of partitions — of numbers, of sets, and of geometric objects — are considered. In fact, the idea of partition can be said to be a leading theme in this book.

The study of partition enumeration was begun by Euler and is very active to this day. We will exposit some parts of this theory. All along the way there are interesting connections with algebra, but these are unfortunately too sophisticated to go into details here. We will, however, give a few hints of this connection, especially in Chapter 5. We also illustrate (in Chapter 15) the relevance of partitions to applied problems.

Another, more recent, topic within enumeration is to count the number of *tilings*. These are partitions of a geometric region into smaller regions of some specified kinds. We will give some glimpses of recent progress in this area. The mathematical roots are in this case mainly from statistical mechanics.

In Chapter 12 we present some progress made in the combinatorial study of convex polytopes. In three dimensions these are the decorative solid bodies with flat polygon sides (such as pyramids, cubes and geodesic domes) that have charmed and intrigued mathematicians and laymen alike since antiquity. In higher dimensions they can be perceived only via mathematical tools, but they are just as beautiful and fascinating. Of this huge subject we discuss the question of laws governing the numbers of faces of various dimensions on the boundary of a polytope.

Combinatorics is used in many ways in computer science, for instance for the construction and analysis of various algorithms. (Remark : *algorithms* are the logically structured systems of commands that instruct computers how to perform prescribed tasks.) Of this young but already huge and rapidly growing area we will give here but the smallest glimpse, namely a couple of examples from complexity theory. This is the part of theoretical computer science that concerns itself with questions about computer calculations of the type "How hard is it ?", "How much time will it take ?" Proving that you cannot do better than what presently known methods allow is often the hardest part, and the part where the most mathematics is needed. Our examples are of this kind.

To illustrate the surprising connections that exist between combinatorics and seemingly unrelated parts of mathematics we have chosen the links with topology. This is an area which on first acquaintance seems far removed from combinatorics, having to do with very general infinite spaces. Nevertheless, the tools of algebraic topology have proven to be of use for solving some problems from combinatorics and theoretical computer science. Again, the theme of enumeration in its various forms pervades some of this border territory.

Understanding this book should for the most part require no more than some basic knowledge of mathematical notation and concepts involving sets, functions, etc., such as taught in a course on precalculus. Some parts should

be accessible to readers with even less background knowledge while others are more demanding, at least in some of the details. Generally speaking, we start out at a very elementary level and assume more and more mathematical background as we go along. We hope that this way the book is informative for laymen as well as for students and colleagues from other parts of mathematics.

1
Bijective proofs

We mentioned in the introduction that enumeration is a basic aspect of combinatorics. The fundamental problem of enumeration is determining the number $\#S$ of elements of a finite set S. Usually there will be infinitely many finite sets, say $S_1, S_2, \ldots$, and we want to determine the number of elements of all the sets S_n. There is no precise definition of what is meant by a "determination" of the number of elements of the S_n's. Generally speaking, an adequate determination should involve a method for computing each $\#S_n$ that involves considerably less effort than a "brute force" listing of the elements of S_n. For instance, an explicit formula such as $\#S_n = 2^n$ is certainly a nice determination. Another way to determine the $\#S_n$'s is to give a simple *generating function* for these numbers. This method is illustrated in Chapters 2 and 3.

BIJECTIONS

Perhaps the nicest way to determine $\#S$ is to find another set T whose number $\#T$ of elements is known, and then to give a one-to-one correspondence or *bijection* $\varphi : S \to T$. This means that φ is a function from S to T such that (1) if $\varphi(a) = \varphi(b)$ then $a = b$, and (2) for each $t \in T$ there is some $a \in S$ for which $\varphi(a) = t$. It then follows that φ has an *inverse* $\varphi^{-1} : T \to S$, given by $\varphi^{-1}(t) = a$ if $\varphi(a) = t$, which is also a bijection. Since the bijection

φ "pairs up" the elements of S with those of T, it follows that $\#S = \#T$. This method of determining $\#S$ is known as a *bijective proof*. All information we know about T has been "transferred" to S, so in principle we understand $\#S$ as well as we do $\#T$.

In subsequent chapters we will be giving some rather intricate bijective proofs. In many cases we will just define the bijection $\varphi : S \to T$ and omit the actual proof, which could be quite difficult, that φ is indeed a bijection. In the remainder of this chapter we will give a few relatively simple examples of bijective proofs as preparation for the more complex bijections yet to come.

The basis for all bijective proofs is the combinatorial or set-theoretic significance of addition and multiplication. If S and T disjoint finite sets, then

$$\#(S \cup T) = \#S + \#T.$$

The *cartesian product* of S and T is given by

$$S \times T = \{(s,t) \,:\, s \in S,\ t \in T\},$$

the set of all ordered pairs whose first coordinate lies in S and second coordinate in T. By the definition of multiplication as repeated addition it follows that

$$\#(S \times T) = (\#S)(\#T).$$

This reasoning extends to a cartesian product of any finite number of finite sets:

$$\begin{aligned} S_1 \times S_2 \times \cdots \times S_k &= \{(s_1, s_2, \ldots, s_k) \,:\, s_i \in S_i\} \\ \#(S_1 \times S_2 \times \cdots \times S_k) &= (\#S_1)(\#S_2)\cdots(\#S_k). \end{aligned} \tag{1.1}$$

SUBSETS

Now let us consider a fundamental problem of elementary enumeration, viz., determining the number $f(n)$ of subsets of an n-element set, say $\{1, 2, \ldots, n\}$. If we were not particularly inspired, we could argue as follows. Let S be a subset of $\{1, 2, \ldots, n\}$, where $n \geq 1$. If $n \in S$, then there are $f(n-1)$ possibilities for the remaining elements of S (since they can form any subset of $\{1, 2, \ldots, n-1\}$). Similarly if $n \notin S$ there are $f(n-1)$ possibilities for the rest of S. Hence

$$f(n) = 2f(n-1), \quad n \geq 1. \tag{1.2}$$

This is a *recurrence* for $f(n)$. We also have the *initial condition* $f(0) = 1$, since the 0-element set $\emptyset$ (the empty set) has one subset, namely, itself. It is clear that the recurrence (1.2), together with the initial condition $f(0) = 1$, has a unique solution; we first obtain $f(1) = 2 \cdot f(0) = 2 \cdot 1 = 2$, then $f(2) = 2 \cdot f(1) = 2 \cdot 2 = 4$, etc. Moreover, the recurrence (1.2) is so simple that it is easy to obtain "by inspection" a formula for the solution: $f(n) = 2^n$. Even if this answer was only guessed to be correct, it is a simple matter to check that it satisfies $f(0) = 1$ and the recurrence (1.2). This technique of finding and solving recurrences is very common in enumerative combinatorics, and many sophisticated techniques have been developed for solving recurrences.

Since the number of subsets of $\{1, 2, \ldots, n\}$ turned out to be $f(n) = 2^n = 2 \times 2 \times \cdots \times 2$ (n times), we can ask if there is a more direct way to see it. In other words, is there a simple bijective proof that $f(n) = 2^n$? Here the problem is so simple that the bijection is quite transparent. Let $\mathcal{S}_n$ denote the set of all subsets of $\{1, 2, \ldots, n\}$. For instance,

$$\mathcal{S}_3 = \{\emptyset, 1, 2, 3, 12, 13, 23, 123\},$$

where we abbreviate a subset such as $\{1, 3\}$ by 13. Similarly let

$$\mathcal{T}_n = \{(a_1, a_2, \ldots, a_n) : a_i = 0 \text{ or } 1\},$$

the set of all binary n-tuples. For instance,

$$\mathcal{T}_3 = \{000, 100, 010, 001, 110, 101, 011, 111\}.$$

Define a function $\varphi : \mathcal{S}_n \to \mathcal{T}_n$ as follows: if S is a subset of $\{1, 2, \ldots, n\}$ then set $\varphi(S) = (a_1, a_2, \ldots, a_n)$, where

$$a_i = \begin{cases} 1, & i \in S \\ 0, & i \notin S. \end{cases}$$

For instance, if $n = 8$ then

$$\varphi(\{2, 3, 6\}) = (0, 1, 1, 0, 0, 1, 0, 0).$$

It should be clear that φ is a bijection. In other words, for every sequence $\alpha = (a_1, a_2, \ldots, a_n) \in \mathcal{T}_n$, there is a unique $S \in \mathcal{S}_n$ such that $\varphi(S) = \alpha$. In fact, we simply define $i \in S$ if $a_i = 1$ and $i \notin S$ if $a_i = 0$. Since $\#\mathcal{T}_n = 2^n$ by equation (1.1), we conclude that $\#\mathcal{S}_n = 2^n$.

COMPOSITIONS

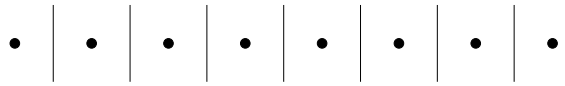
Let us consider some less obvious bijective proofs. A *composition* of an integer $n \geq 1$ is a way of writing n as an ordered sum of positive integers,

i.e., $n = a_1 + a_2 + \cdots + a_k$, where each a_i is a positive integer. (We allow $k = 1$, i.e., a sum with the single term n.) The terminology "ordered sum" means we take into account the order of the summands. Thus $2 + 1$ is a different composition from $1 + 2$. For instance, there are eight compositions of 4, namely,

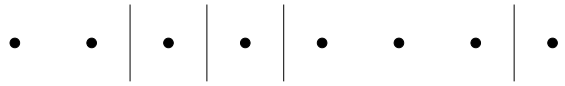
$$1+1+1+1, \quad 2+1+1, \quad 1+2+1, \quad 1+1+2, \quad 1+3, \quad 3+1, \quad 2+2, \quad 4.$$

Let $c(n)$ denote the number of compositions of n. A little computation reveals that $c(1) = 1$, $c(2) = 2$, $c(3) = 4$, $c(4) = 8$, $c(5) = 16$, and $c(6) = 32$, suggesting that $c(n) = 2^{n-1}$. Once this result is guessed, there are numerous ways to prove it. For instance, the number of compositions α of n whose first part is 1 is just $c(n-1)$ (simply remove the first part from α to obtain a composition of $n-1$), while the number of compositions β of n whose first part is greater than 1 is also $c(n-1)$ (simply subtract 1 from the first part of β to obtain a composition of $n-1$). We thus obtain the recurrence $c(n) = 2c(n-1)$, which together with the initial condition $c(1) = 1$ yields $c(n) = 2^{n-1}$.

Since we already know that 2^{n-1} is the number of subsets of an $(n-1)$-element set, we can ask whether there is a bijection between the subsets of an $(n-1)$-element set and the compositions of n. In fact, there is a very elegant such bijection. Consider n dots on a horizontal line. There are $n-1$ spaces between the dots, shown in the illustration below as vertical lines (where $n = 8$).

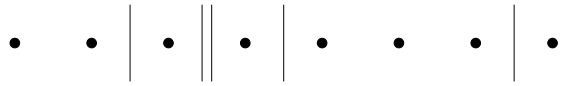


We can choose a subset S of these $n-1$ lines in 2^{n-1} ways. When we retain only the lines in S, these lines divide the n dots into "compartments."



Reading the number of dots in each compartment from left-to-right yields a composition of n. For the example above, we obtain the composition $2+1+1+3+1$ of 8. It should be clear that this argument is bijective, that is, any composition of n corresponds to a unique way of choosing a subset of the spaces between the dots. (If the composition is $a_1 + a_2 + \cdots + a_k$, then choose the space after the first a_1 dots, then after the next a_2 dots, etc.) Note how simple and elegant is this method of representing compositions. It makes their enumeration transparent and shows that compositions are just thinly disguised versions of sets.

We next consider a variation of the previous argument which is not quite so transparent. Let $g(n)$ be the total number of parts that occur in all the compositions of n. For instance, $g(3) = 8$; the relevant compositions, with their number of parts shown in parentheses, are $1+1+1$ (3), $2+1$ (2), $1+2$ (2), and 3 (1). A first attempt to find a formula for $g(n)$ by a bijective argument might be to first choose a composition α of n, represented by dots and vertical lines as above, and then choose a compartment (which corresponds to one of the parts of α). The total number of such choices is $g(n)$. The problem with this approach is that the number of choices for which compartment to choose depends on the choice of α. The trick, in a sense, is to choose the compartment first and then the composition α! Namely, first draw one double line between two of the dots or at the left of all the dots. Then choose a subset of the remaining spaces between the dots, represented by vertical lines.



The compartments formed by the double line and the single lines define a composition α as before. Moreover, the compartment immediately to the right of the double line corresponds to one of the parts of α. Thus the total number of choices is $g(n)$. In the example above α is given by $2+1+1+3+1$, and we have chosen the third out of the five terms of α.

We don't quite have complete independence of the number of choices of single vertical lines from the choice of the double line, but there are only two different cases. If the double line is chosen at the beginning (in one way), then there are 2^{n-1} choices for the single lines. If the double line lies between two dots (in $(n-1)$ ways) then there are 2^{n-2} choices for the single lines. Hence

$$g(n) = 2^{n-1} + (n-1)2^{n-2} = (n+1)2^{n-2}.$$

As in our previous examples the bijective proof, once it is understood, makes the final answer almost obvious.

Subsets with repetition allowed

We will give two further examples of simple bijective proofs. Most readers are probably familiar with the *binomial coefficients* $\binom{n}{k}$. If k, n are nonnegative integers then $\binom{n}{k}$ has a standard combinatorial interpretation: it is the number of k-element subsets of an n-element set. For instance $\binom{4}{2} = 6$, since the subsets of $\{1,2,3,4\}$ are (abbreviating $\{a,b\}$ as ab) 12, 13, 23, 14, 24, 34.

There is also a simple formula for $\binom{n}{k}$, namely,

$$\binom{n}{k} = \begin{cases} \dfrac{n!}{k!(n-k)!}, & \text{if } 0 \le k \le n \\ 0, & \text{if } 0 \le n < k, \end{cases}$$

where $n!$ (read "n factorial") is short for $1 \cdot 2 \cdot \cdots \cdot n$. However, this formula will not be relevant to us here.

Now define $\left(\!\binom{n}{k}\!\right)$ to be the number of ways to choose k elements from an n-element set, disregarding order, if repetitions are allowed. For instance, $\left(\!\binom{3}{2}\!\right) = 6$, since the choices of two elements from $\{1,2,3\}$, allowing repetitions, are 11, 22, 33, 12, 13, 23. Similarly $\left(\!\binom{2}{3}\!\right) = 4$, the choices being 111, 112, 122, 222.

There is a nice trick, part of the standard repertoire of enumerative combinatorics, to reduce the problem of subsets with repetition to those without repetition. Suppose that the n-element set is $\{1,2,\ldots,n\}$. Choose k elements with repetition allowed, and arrange these elements in increasing order:

$$1 \le a_1 \le a_2 \le \cdots \le a_k \le n. \tag{1.3}$$

Let $b_i = a_i + i - 1$. The effect of adding $i-1$ to a_i is to "stretch out" the sequence $a_1,\ldots,a_k$ so that the elements become distinct. Since a_1 is kept the same and a_k is increased by $k-1$, the inequalities $1 \le a_1$ and $a_k \le n$ become $1 \le b_1$ and $b_k \le n+k-1$. Hence we obtain

$$1 \le b_1 < b_2 < \cdots < b_k \le n+k-1, \tag{1.4}$$

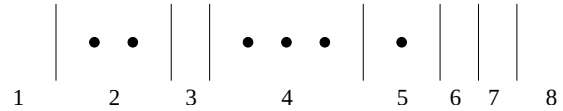
so $b_1, b_2, \ldots, b_k$ form a k-element *subset* of $\{1,2,\ldots,n+k-1\}$ (no repetitions). We can easily reverse this process. Given a k-element subset of $\{1,2,\ldots,n+k-1\}$, write its elements in increasing order as in equation (1.4). Define $a_i = b_i - i + 1$. Then the a_i's satisfy equation (1.3). Hence the correspondence between k-element subsets $\{a_1,\ldots,a_k\}$ of $\{1,2,\ldots,n\}$ with repetition allowed and ordinary k-element subsets $\{b_1,\ldots,b_k\}$ (no repetition) of $\{1,2,\ldots,n+k-1\}$ is a bijection. There follows

$$\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k}.$$

We have therefore solved the problem of counting subsets with repetition by establishing a simple bijection between such subsets and those without repetition.

There is another elegant bijective proof that $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k}$, based on the "dots and slots" method we used to prove that there are 2^{n-1} compositions of n. In our previous use of dots and slots, we first placed the dots and then the vertical bars between them to define the slots. For our present purposes we instead place the dots and bars *together*. More specifically, suppose we have k dots and $n-1$ bars. We write them one after another in *any* order. For instance, suppose we have six dots and seven bars. One way of arranging them is as follows:

The $n-1$ bars form n compartments (slots), including the compartments before the first bar and after the last. For definiteness, label these compartments $1, 2 \ldots, n$ from left to right:



Each of these compartments has some number (which may be 0) of dots in it. For the example above, compartment 2 has two dots, compartment 4 has three dots, and compartment 5 has one dot. All the other compartments have no dots. This placement thus corresponds to choosing 2 twice, 4 three times, and 5 once from the set $\{1,2,3,4,5,6,7,8\}$. Any choice of k elements from $1,\ldots,n$ with repetition allowed can be depicted in this way as an ordering of k dots and $n-1$ bars. Hence the number $\left(\!\binom{n}{k}\!\right)$ of such choices is equal to the number of ways to order k dots and $n-1$ bars. We have an ordering $a_1, a_2, \ldots, a_{n+k-1}$ of $n+k-1$ objects in all, of which k of them are dots. In other words, choose k of the a_i's to be dots, and the remaining $n-1$ to be bars. Hence the number of possible orderings is the number of ways to choose k objects from $n+k-1$, which is just the binomial coefficient $\binom{n+k-1}{k}$. We have therefore given a second bijective proof that $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k}$.

LABELED TREES

Combinatorics abounds with bijective proofs, many of which are beautiful and instructive. Some are also quite surprising. Let us present one of the classics.

By a *tree* we understand a collection of n objects, called *nodes*, and $n-1$ unordered pairs of nodes, called *edges*, such that the edges connect the nodes. A picture shows the idea; see Figure 1.

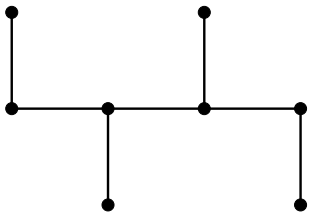


FIGURE 1
A tree.

For instance, we can think of the nodes as towns and the edges as fiber cable links connecting certain pairs of towns, in such a way that the resulting network connects all towns with as few links as possible.

A *labeled tree* is a tree with names attached to the nodes, distinct names for distinct nodes. If the tree has n nodes we can take as a standard set of labels the integers $1, 2, \ldots, n$. Figure 2 shows a labeled tree on eight nodes.

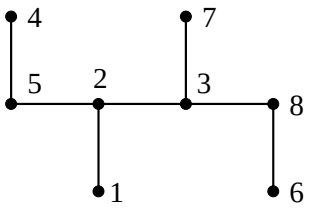


FIGURE 2
A labeled tree.

Two labeled trees are considered to be distinct if for some pair of numbers i and j the corresponding nodes are connected by an edge in one tree and not in the other. We now ask,

how many distinct labeled trees are there on n nodes?

The answer was found by the British mathematician James Joseph Sylvester in 1857, though he did not publish a proof. The first proof was given by the German mathematician Carl Wilhelm Borchardt in 1860. The result is often attributed to the British mathematician Arthur Cayley and is sometimes called *Cayley's theorem*. Cayley was the first mathematician to systematically investigate trees, which he did while studying the structure of chemical molecules.

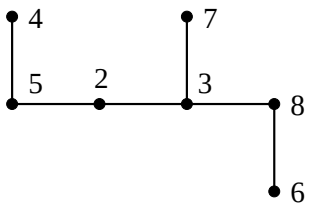
Cayley's Theorem. *The number of labeled trees on n nodes is n^{n-2}.*

The theorem reveals an instance of what is called "combinatorial explosion". This refers to the stunning growth of the number of possibilities in many seemingly simple combinatorial situations. For instance, the theorem tells us that there are 100 million ways to minimally connect 10 cities by fiber cables, in the manner described above.

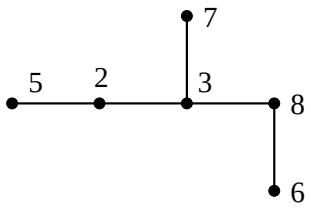
A bijective proof of Cayley's theorem was given in 1918 by Ernst Paul Heinz Prüfer . This proof sets up a bijection between the set of labeled trees on n nodes and the set of strings $(a_1, \ldots, a_{n-2})$ where each of the entries a_i is one of the numbers $1, 2, \ldots, n$. The number of such strings is by equation (1.1) equal to n^{n-2}.

Here is a description of the bijection. Let T be a labeled tree. Find the leaf (node with only one neighbor) with the lowest label. Write down the label of its neighbor, and then delete that leaf. Then just repeat this simple step until a string of length $n-2$ is obtained.

For instance, letting T be the tree in Figure 2, the "smallest" leaf is the one labeled by 1. Hence we record the label of its neighbor, namely, 2, and delete that leaf. The following labeled tree then remains:



The smallest leaf is now the one labeled by 4, so we record 5 and delete. We then have:



Here the smallest leaf is the one labeled by 5, so we record 2 and delete. And so on, ... After 6 steps we have recorded the string $(2,5,2,3,8,3)$, and we are done.

That the string $(2,5,2,3,8,3)$ represents the tree of Figure 2 under Prüfer's bijection means that only knowing this string we can fully reconstruct the labeled tree. We leave to the reader to figure out how to construct the tree corresponding to a string in general. As a hint, note that the smallest missing label from the string $(2,5,2,3,8,3)$, namely 1, is the label of the first leaf removed. Hence 1 and 2 are connected by an edge. This is the first step in reconstructing the tree.

We hope that the examples of this chapter have given the reader an idea of what is meant by a bijective proof and why such proofs are enlightening. In subsequent chapters we will encounter some much more complicated bijections.

2
Partitions

A fundamental concept in combinatorics is that of a partition. In general, a partition of an object is a way of breaking it up into smaller objects. We will be concerned here with partitions of *positive integers* (positive whole numbers). Later on we will encounter also other kinds of partitions. The subject of partitions has a long history going back to Gottfried Wilhelm von Leibniz (1646–1716) and Euler, and has been found to have unexpected connections with a number of other subjects.

NUMBER PARTITIONS

A *partition* of a positive integer n is a way of writing n as a sum of positive integers, ignoring the order of the summands. For instance, $3+4+2+1+1+4$ represents a partition of 15, and $4+4+3+2+1+1$ represents the same partition. We allow a partition to have only one part (summand), so that 5 is a partition of 5. There are in fact seven partitions of 5, given by

$$
\begin{aligned}
&5\\
&4+1\\
&3+2\\
&3+1+1\\
&2+2+1\\
&2+1+1+1\\
&1+1+1+1+1.
\end{aligned}
$$

Contrast the definition of a partition with that of a composition (as defined in Chapter 1), where the order of the summands does matter.

We denote the number of partitions of n by $p(n)$, so for instance $p(5) = 7$. By convention we set $p(0) = 1$, and similarly for related partition functions discussed below. The problem of evaluating $p(n)$ has a long history. There is no simple formula in general for $p(n)$, but there are remarkable and quite sophisticated methods to compute $p(n)$ for "reasonable" values of n. For instance, as long ago as 1938 Derrick Henry Lehmer computed $p(14,031)$ (a number with 127 decimal digits!), and nowadays a computer would have no trouble computing $p(10^{15})$, a number with 35,228,031 decimal digits. In general, a good approximation to the number of decimal digits of $p(n)$ is given by

$$1.1140086280105007831\cdots\sqrt{n}.$$

For $n = 10^{15}$ this approximation is (to the nearest integer) 35,228,046. For those familiar with the natural logarithm $\log(x)$, let us remark that the constant $1.114008\cdots$ appearing above is equal to $\pi\sqrt{2/3}/\log(10)$, a consequence of a famous "asymptotic formula" for $p(n)$ due to Godfrey Harold Hardy and Srinivasa Aiyangar Ramanujan in 1918. This formula gives not just a good approximation to the number of decimal digits of $p(n)$, but in fact a good approximation to $p(n)$ itself.

Generating functions

It is also possible to codify all the numbers $p(n)$ into a single object known as a *generating function*. A generating function (in the variable x) is an expression of the form

$$F(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots,$$

where the coefficients $a_0, a_1, \ldots$ are numbers. (We call a_n the *coefficient* of x^n, and call a_0 the *constant term*. The notation x^0 next to a_0 is suppressed.) The generating function $F(x)$ differs from a polynomial in x in that it can have infinitely many terms. We regard x as a formal symbol, and do not think

of it as standing for some unknown quantity. Thus the generating function $F(x)$ is just a way to represent the sequence $a_0, a_1, \ldots$.

It is natural to ask what advantage is gained in representing a sequence in such a way. The answer is that generating functions can be manipulated in various ways that often are useful for combinatorial problems. For instance, letting $G(x) = b_0 + b_1x + b_2x^2 + \cdots$, we can add $F(x)$ and $G(x)$ by the rule

$$F(x) + G(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots.$$

In other words, we simply add the coefficients, just as we would expect from the ordinary rules of algebra. Similarly we can form the product $F(x)G(x)$ using the ordinary rules of algebra, in particular the law of exponents $x^i x^j = x^{i+j}$. To perform this multiplication, we pick a term a_ix^i from $F(x)$ and a term b_jx^j from $G(x)$ and multiply them to get $a_ib_jx^{i+j}$. We then add together all such terms. For instance, the term in the product involving x^4 will be

$$\begin{aligned} a_0 \cdot b_4x^4 + a_1x \cdot b_3x^3 + a_2x^2 \cdot b_2x^2 + a_3x^3 \cdot b_1x + a_4x^4 \cdot b_0 \\ = (a_0b_4 + a_1b_3 + a_2b_2 + a_3b_1 + a_4b_0)x^4. \end{aligned}$$

In general, the coefficient of x^n in $F(x)G(x)$ will be

$$a_0b_n + a_1b_{n-1} + a_2b_{n-2} + \cdots + a_{n-1}b_1 + a_nb_0.$$

Consider for instance the product of $F(x) = 1 + x + x^2 + x^3 + \cdots$ with $G(x) = 1 - x$. The constant term is just $a_0b_0 = 1 \cdot 1 = 1$. If $n > 1$ then the coefficient of x^n is $a_nb_0 + a_{n-1}b_1 = 1 - 1 = 0$ (since $b_i = 0$ for $i > 1$, so we have only two nonzero terms). Hence

$$(1 + x + x^2 + x^3 + \cdots)(1 - x) = 1.$$

For this reason we write

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots.$$

Some readers will recognize this formula as the sum of an infinite geometric series, though here the formula is "formal," that is, x is regarded as just a symbol and there is no question of convergence. Similarly, for any $k \geq 1$ we get

$$\frac{1}{1-x^k} = 1 + x^k + x^{2k} + x^{3k} + \cdots. \tag{2.1}$$

Now let $P(x)$ denote the (infinite) product

$$P(x) = \frac{1}{1-x} \cdot \frac{1}{1-x^2} \cdot \frac{1}{1-x^3} \cdots.$$

We may also write this product as

$$P(x) = \frac{1}{(1-x)(1-x^2)(1-x^3)\cdots}. \tag{2.2}$$

Can any sense be made of this product? According to our previous discussion, we can rewrite the right-hand side of equation (2.2) as

$$P(x) = (1 + x + x^2 + \cdots)(1 + x^2 + x^4 + \cdots)(1 + x^3 + x^6 + \cdots)\cdots .$$

To expand this product as a sum of individual terms, we must pick a term x^{m_1} from the first factor, a term x^{2m_2} from the second, a term x^{3m_3} from the third, etc., multiply together all these terms, and then add all such products together. In order not to obtain an infinite (and therefore meaningless) exponent of x, it is necessary to stipulate that when we pick the terms $x^{m_1}, x^{2m_2}, x^{3m_3}, \ldots$, only finitely many of these terms are not equal to 1. (Equivalently, only finitely many of the m_i are not equal to 0.) We then obtain a single term $x^{m_1+2m_2+3m_3+\cdots}$, where the exponent $m_1 + 2m_2 + 3m_3 + \cdots$ is finite. The coefficient of x^n in $P(x)$ will then be the number of ways to write n in the form $m_1 + 2m_2 + 3m_3 + \cdots$ for nonnegative integers $m_1, m_2, m_3, \ldots$. But writing n in this form is the same as writing n as a sum of m_1 1's, m_2 2's, m_3 3's, etc. Such a way of writing n is just a partition of n. For instance, the partition $5 + 5 + 5 + 4 + 2 + 2 + 2 + 2 + 1 + 1 + 1$ of 30 corresponds to choosing $m_1 = 3, m_2 = 4, m_4 = 1, m_5 = 3$, and all other $m_i = 0$. It follows that the coefficient of x^n in $P(x)$ is just $p(n)$, the number of partitions of n, so we obtain the famous formula of Euler

$$p(0) + p(1)x + p(2)x^2 + \cdots = \frac{1}{(1-x)(1-x^2)(1-x^3)\cdots}. \tag{2.3}$$

Odd parts and distinct parts

Although equation (2.3) is very elegant, one may ask whether it is of any use. Can it be used to obtain interesting information about the numbers $p(n)$? To answer that, let us show how simple manipulation of generating functions (due to Euler) gives a surprising connection between two types of partitions. Let $r(n)$ be the number of partitions of n into *odd* parts. For instance, $r(7) = 5$, the relevant partitions being

$$7 = 5 + 1 + 1 = 3 + 3 + 1 = 3 + 1 + 1 + 1 + 1 = 1 + 1 + 1 + 1 + 1 + 1 + 1.$$

Let

$$R(x) = r(0) + r(1)x + r(2)x^2 + r(3)x^3 + \cdots .$$

Exactly as equation (2.3) was obtained we get

$$R(x) = \frac{1}{(1-x)(1-x^3)(1-x^5)(1-x^7)\cdots}. \tag{2.4}$$

Similarly, let $q(n)$ be the number of partitions of n into *distinct* parts, that is, no integer can occur more than once as a part. For instance, $q(7) = 5$, the relevant partitions being

$$7 = 6+1 = 5+2 = 4+3 = 4+2+1.$$

Note that $r(7) = q(7)$. In order to explain this "coincidence," let

$$Q(x) = q(0) + q(1)x + q(2)x^2 + q(3)x^3 + \cdots .$$

The reader who understands the derivation of equation (2.3) will have no trouble seeing that

$$Q(x) = (1+x)(1+x^2)(1+x^3)\cdots . \tag{2.5}$$

Now we come to the ingenious trick of Euler. Note that by ordinary "high school algebra," we have

$$1 + x^n = \frac{1-x^{2n}}{1-x^n}.$$

Thus from equation (2.5) we obtain

$$\begin{aligned} Q(x) &= \frac{1-x^2}{1-x}\cdot\frac{1-x^4}{1-x^2}\cdot\frac{1-x^6}{1-x^3}\cdot\frac{1-x^8}{1-x^4}\cdots \\ &= \frac{(1-x^2)(1-x^4)(1-x^6)(1-x^8)\cdots}{(1-x)(1-x^2)(1-x^3)(1-x^4)\cdots}. \end{aligned} \tag{2.6}$$

When we cancel the factors $1-x^{2i}$ from both the numerator and denominator, we are left with

$$Q(x) = \frac{1}{(1-x)(1-x^3)(1-x^5)\cdots},$$

which is just the product formula (2.4) for $R(x)$. This means that $Q(x) = R(x)$. Thus the coefficients of $Q(x)$ and $R(x)$ are the same, so we have proved that $q(n) = r(n)$ for all n. In other words, we have the following result.

Theorem (Euler). *For every n the number of partitions of n into distinct parts equals the number of partitions of n into odd parts.*

The above argument shows the usefulness of working with generating functions. Many similar generating function techniques have been developed that make generating functions into a fundamental tool of enumerative combinatorics.

Once we obtain a formula such as $q(n) = r(n)$ by an indirect means like generating functions, it is natural to ask whether there might be a simpler proof. For the problem at hand, we would like to correspond to each partition of n into distinct parts a partition of n into odd parts, such that every partition of n into odd parts is associated with exactly one partition of n into distinct parts, and conversely every partition of n into distinct parts is associated with exactly one partition of n into odd parts. In other words, we want a *bijective proof* of the formula $q(n) = r(n)$. Several such proofs are known; we give the perhaps simplest of these, due to James Whitbread Lee Glaisher. It is based on the fact that every positive integer n can be uniquely written as a sum of distinct powers of two — this is simply the binary expansion of n. For instance, $10000 = 2^{13} + 2^{10} + 2^9 + 2^8 + 2^4$. Suppose we are given a partition into odd parts, such as

$$202 = 19 + 19 + 19 + 11 + 11 + 11 + 11 + 9 + 7 + 7 + 7 + 5$$

$$+5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 1 + 1 + 1 + 1 + 1 + 1.$$

We can rewrite this partition as

$$3 \cdot 19 + 4 \cdot 11 + 1 \cdot 9 + 3 \cdot 7 + 13 \cdot 5 + 6 \cdot 1,$$

where each part is multiplied by the number of times it appears. This is just the expression $m_1 + 2m_2 + 3m_3 + \cdots$ for a partition discussed above. Now write each of the numbers m_i as a sum of distinct powers of 2. For the above example, we get

$$202 = (2 + 1) \cdot 19 + 4 \cdot 11 + 1 \cdot 9 + (2 + 1) \cdot 7 + (8 + 4 + 1) \cdot 5 + (4 + 2) \cdot 1.$$

Expand each product into a sum (by the distributivity of multiplication over addition):

$$202 = (38 + 19) + 44 + 9 + (14 + 7) + (40 + 20 + 5) + (4 + 2). \tag{2.7}$$

We have produced a partition of the same number n with distinct parts. That the parts are distinct is a consequence of the fact that every integer n can be uniquely written as the product of an odd number and a power of 2 (keep on dividing n by 2 until an odd number remains). Moreover, the whole procedure can be reversed. That is, given a partition into distinct parts such as

$$202 = 44 + 40 + 38 + 20 + 19 + 14 + 9 + 7 + 5 + 4 + 2,$$

group the terms together according to their largest odd divisor. For instance, 40, 20, and 5 have the largest odd divisor 5, so we group them together. We

thus recover the grouping (2.7). We can now factor the largest odd divisor d out of each group, and what remains is the number of times d appears as a part. Thus we have recovered the original partition. This reasoning shows that we have indeed produced a bijection between partitions of n into odd parts and partitions of n into distinct parts. It provides a "natural" explanation of the fact that $q(n) = r(n)$, unlike the generating function proof which depended on a miraculous trick.

ROGERS-RAMANUJAN AND BEYOND

The subject of partitions is replete with results similar to Euler's, in which two sets of partitions turn out to have the same number of elements. The most famous of these results is called the *Rogers-Ramanujan identities*, after Leonard James Rogers and Ramanujan, who in 1894 and 1913 proved these identities in the form of an identity between generating functions. It was Percy Alexander MacMahon who interpreted them combinatorially as follows.

First Rogers-Ramanujan Identity. *Let $f(n)$ be the number of partitions of n whose parts differ by at least 2. For instance, $f(13) = 10$, the relevant partitions being*

$$13 = 12 + 1 = 11 + 2 = 10 + 3 = 9 + 4 = 8 + 5 = 9 + 3 + 1$$
$$= 8 + 4 + 1 = 7 + 5 + 1 = 7 + 4 + 2.$$

Similarly, let $g(n)$ be the number of partitions of n whose parts are of the form $5k + 1$ or $5k + 4$ (i.e., leave a remainder of 1 or 4 upon division by 5). For instance, $g(13) = 10$:

$$11 + 1 + 1 = 9 + 4 = 9 + 1 + 1 + 1 + 1 = 6 + 6 + 1 = 6 + 4 + 1 + 1 + 1$$
$$= 6+1+1+1+1+1+1 = 4+4+4+1 = 4+4+1+1+1+1+1$$
$$= 4+1+1+1+1+1+1+1+1 = 1+1+1+1+1+1+1+1+1+1+1+1+1.$$

Then $f(n) = g(n)$ for every n.

Second Rogers-Ramanujan Identity. *Let $u(n)$ be the number of partitions of n whose parts differ by at least 2 and such that 1 is not a part. For instance, $u(13) = 6$, the relevant partitions being*

$$13 = 11 + 2 = 10 + 3 = 9 + 4 = 8 + 5 = 7 + 4 + 2.$$

Similarly, let $v(n)$ be the number of partitions of n whose parts are of the form $5k + 2$ or $5k + 3$ (i.e., leave a remainder of 2 or 3 upon division by 5). For instance, $v(13) = 6$:

$$13 = 8+3+2 = 7+3+3 = 7+2+2+2 = 3+3+3+2+2 = 3+2+2+2+2+2.$$

Then $u(n) = v(n)$ *for every* n.

The Rogers-Ramanujan identities have been given many proofs, but none of them is really easy. The important role played by the number 5 seems particularly mysterious. The first bijective proof of the Rogers-Ramanujan identities was given by Issai Schur in 1917. An interesting later bijective proof is due in 1980 to Adriano Mario Garsia and Stephen Carl Milne, as a special case of a general scheme for finding bijective proofs of partition identities. However, both these proofs are rather complicated and involve cancellation arguments. It would be greatly desirable to find a simpler, more direct bijective proof, though a 2004 paper by Igor Pak shows that a certain type of simple proof does not exist.

The Rogers-Ramanujan identities and related identities are not just number-theoretic curiosities. They have arisen completely independently in several seemingly unrelated areas. To give just one example, a famous open problem in statistical mechanics, known as the *hard hexagon model*, was solved in 1980 by Rodney James Baxter using the Rogers-Ramanujan identities.

The subject of partition identities has received so much attention since Euler that one would not expect a whole new class of relatively simple identities to have remain undiscovered until recently. However, just such a class of identities was found by Mireille Bousquet-Mélou and Kimmo Eriksson beginning in 1996. We will state one of the simplest of their identities to give the reader the striking flavor of their results.

The *Lucas numbers* L_n are defined by the conditions $L_1 = 1$, $L_2 = 3$, and $L_{n+1} = L_n + L_{n-1}$ for $n \geq 2$. Thus $L_3 = 4$, $L_4 = 7$, $L_5 = 11$, $L_6 = 18$, $L_7 = 29$, etc. Those familiar with Fibonacci numbers will see that the Lucas numbers satisfy the same recurrence as Fibonacci numbers, but with the initial conditions $L_1 = 1$ and $L_2 = 3$, rather that $F_1 = F_2 = 1$ for Fibonacci numbers. Let $f(n)$ be the number of partitions of n all of whose parts are Lucas numbers L_{2n+1} of odd index. For instance, we have $f(12) = 5$, corresponding to the partitions

$$\begin{array}{l} 1+1+1+1+1+1+1+1+1+1+1+1 \\ 4+1+1+1+1+1+1+1+1 \\ 4+4+1+1+1+1 \\ 4+4+4 \\ 11+1 \end{array} .$$

Let $g(n)$ be the number of partitions of n into parts $a_1 \le a_2 \le \cdots \le a_k$ such that $a_i/a_{i-1} > \frac{1}{2}(3+\sqrt{5}) = 2.618\cdots$ for all i. For instance, $g(12) = 5$, corresponding to the partitions

$$12, \quad 11+1, \quad 10+2, \quad 9+3, \quad \text{and} \quad 8+3+1.$$

Note that the number $\frac{1}{2}(3+\sqrt{5})$ used to define $g(n)$ is the square of the "golden ratio" $\frac{1}{2}(1+\sqrt{5})$. The surprising result of Bousquet-Mélou and Eriksson is that $f(n) = g(n)$ for all n.

3
Plane partitions

A partition such as $8+6+6+5+2+2+2+2+1+1$ may be regarded simply as a linear array of positive integers,

$$8\ 6\ 6\ 5\ 2\ 2\ 2\ 2\ 1\ 1$$

whose entries are *weakly decreasing*, i.e., each entry is greater than or equal to the one on its right. Viewed in this way, one can ask if there are interesting "multidimensional" generalizations of partitions, in which the parts don't lie on just a line, but rather on some higher dimensional object. The simplest generalization occurs when the parts lie in a plane.

COUNTING PLANE PARTITIONS

Rather than having the parts weakly decreasing in a single line, we now want the parts to be weakly decreasing in every row and column. More precisely, let λ be a partition with its parts $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ written in weakly decreasing order, so $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell > 0$. We define a *plane partition* π of *shape* λ to be a left-justified array of positive integers (called the *parts* of π) such that (1) there are λ_i parts in the ith row, and (2) every row (read left-to-right) and column (read top-to-bottom) is weakly decreasing. An example of a plane partition is given in Figure 3.

```
7 4 4 4 2 2 1 1 1 1
7 4 4 2 2 1 1 1 1
6 3 2 2 2 1 1 1 1
4 2 2 1 1 1
2 2 1 1 1
2 1 1 1 1
1 1 1 1 1
1 1
```

FIGURE 3
A plane partition

We say that π is a plane partition *of* n if n is the sum of the parts of π. Thus the plane partition of Figure 3 is a plane partition of 100, of shape $(10, 9, 9, 6, 5, 5, 5, 2)$. It is clear what is meant by the *number of rows* and *number of columns* of π. For the example in Figure 3, the number of rows is 8 and the number of columns is 10. The plane partitions of integers up to 3 (including the empty set $\emptyset$, which is regarded as a plane partition of 0) are given by

$$\emptyset \quad 1 \quad 2 \quad 11 \quad \begin{matrix}1\\1\end{matrix} \quad 3 \quad 21 \quad 111 \quad \begin{matrix}11\\1\end{matrix} \quad \begin{matrix}2\\1\end{matrix} \quad \begin{matrix}1\\1\\1\end{matrix}\ .$$

Thus, for instance, there are six plane partitions of 3.

In 1912 MacMahon began a study of the theory of plane partitions. MacMahon was a mathematician well ahead of his time. He worked in virtual isolation on a variety of topics within enumerative combinatorics that did not become fashionable until many years later. A highlight of MacMahon's work was a simple generating function for the number of plane partitions of n. More precisely, let $pp(n)$ denotes the number of plane partitions of n, so that $pp(0) = 1,\ pp(1) = 1,\ pp(2) = 3,\ pp(3) = 6,\ pp(4) = 13,$ etc.

MacMahon's Theorem.

$$pp(0) + pp(1)x + pp(2)x^2 + pp(3)x^3 + \cdots$$

$$= \frac{1}{(1-x)(1-x^2)^2(1-x^3)^3(1-x^4)^4\cdots}. \tag{3.1}$$

Unlike Euler's formula (2.3) for the generating function for the number $p(n)$ of ordinary partitions of n, MacMahon's remarkable formula is by no means easy

to prove. MacMahon's proof was an intricate induction argument involving manipulations of determinants. Only much later was a bijective proof found by Edward Anton Bender and Donald Ervin Knuth. Their proof was based on the *RSK algorithm*, a central result in enumerative combinatorics and its connections with the branch of mathematics known as *representation theory*. This correspondence was first stated by Gilbert de Beauregard Robinson in a rather vague form in 1938 (with some assistance from Dudley Ernest Littlewood), and later more explicitly by Craige Eugene Schensted in 1961. Schensted's motivation for looking at this correspondence is discussed in Chapter 6. The version of the RSK algorithm used here is due to Knuth. Hence the letters RSK stand for Robinson, Schensted, and Knuth.

The RSK algorithm and the Bender-Knuth proof

We now give a brief account of the proof of Bender and Knuth. Using equation (2.1), the product on the right-hand side of (3.1) may be written

$$\frac{1}{(1-x)(1-x^2)^2(1-x^3)^3(1-x^4)^4\cdots} = (1+x+x^2+\cdots)(1+x^2+x^4+\cdots)$$

$$(3.2)\qquad (1+x^2+x^4+\cdots)(1+x^3+x^6+\cdots)(1+x^3+x^6+\cdots)(1+x^3+x^6+\cdots)\cdots .$$

In general, there will be k factors of the form $1+x^k+x^{2k}+x^{3k}+\cdots$. We must pick a term out of each factor (with only finitely many terms not equal to 1) and multiply them together to get a term x^n of the product. A bijective proof of (3.1) therefore consists of associating a plane partition of n with each choice of terms from the factors $1+x^k+x^{2k}+\cdots$, such that the product of these terms is x^n.

Our first step is to encode a choice of terms from each factor by an array of numbers called a *two-line array*. A typical two-line array A looks like

$$(3.3)\qquad A = \begin{array}{cccccccccccccc} 3 & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 3 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 4 & 4 & 3 & 3 & 3 \end{array} .$$

The first line is a (finite) weakly decreasing sequence of positive integers. The second line consists of a positive integer below each entry in the first line, such that the integers in the second line appearing below equal integers in the first line are in weakly decreasing order. For instance, for the two-line array A above, the integers appearing below the 2's of the first line are 2 2 2 1 1 (in that order). Such a two-line array encodes a choice of terms from the factors of the product (3.2) as follows. Let a_{ij} be the number of columns $\begin{smallmatrix} i \\ j \end{smallmatrix}$ of A. For instance (always referring to the two-line array (3.3)),

$a_{33} = 1,\ a_{31} = 2,\ a_{13} = 3,\ a_{23} = 0$. Given a_{ij}, let $k = i+j-1$. Then choose the term $x^{a_{ij} \cdot k}$ from the ith factor of (3.2) of the form $1 + x^k + x^{2k} + \cdots$. For instance, since $a_{33} = 1$ we have $k = 5$ and choose the term $x^{1 \cdot 5} = x^5$ from the third factor of the form $1 + x^5 + x^{10} + \cdots$. Since $a_{31} = 2$ we have $k = 3$ and choose the term $x^{2 \cdot 3} = x^6$ from the third factor of the form $1 + x^3 + x^6 + \cdots$, etc. In this way we obtain a one-to-one correspondence between a choice of terms from each factor of the product (3.2) (with only finitely terms not equal to 1) and two-line arrays A.

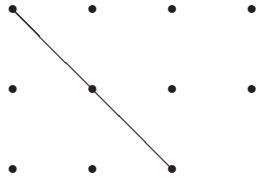
We now describe the part of the Bender-Knuth bijection which is the RSK algorithm. We will insert the numbers in each line of the two-line array A into a successively evolving plane partition, yielding in fact a pair of plane partitions. These plane partitions will have the special property of being *column-strict*, that is, the (nonzero) entries are *strictly* decreasing in each column. Thus after we have inserted the first i numbers of the first and second lines of A, we will have a pair P_i and Q_i of column-strict plane partitions. We insert the numbers of the second line of A successively into P_i from left-to-right by the following rule. Assuming that we have inserted the first $i-1$ numbers, yielding P_{i-1} and Q_{i-1}, we insert the ith number a of the second row of A into P_{i-1}, by putting it as far to the right as possible in the first row of P_{i-1} so that this row remains weakly decreasing. In doing so, it may displace (or *bump*) another number b already in the first row. Then insert b into the second row according to the same rule, that is, as far to the right as possible so that the second row remains weakly decreasing. Then b may bump a number c into the third row, etc. Continue this "bumping procedure" until finally a number is inserted at the end of the row, thereby not bumping another number. This yields the column-strict plane partition P_i. (It takes a little work, which we omit, to show that P_i is indeed column-strict.) Now insert the ith number of the first row of A (that is, the number directly above the a in A that we have just inserted into P_{i-1} to form P_i)) into Q_{i-1} to form Q_i, by placing it so that P_i and Q_i have the same *shape*, that is, the same number of elements in each row. If A has m columns, then the process stops after obtaining P_m and Q_m, which we denote simply as P and Q.

Example. Figure 4 illustrates the bumping procedure with the two-line array A of equation (3.3). The numbers that have bumped other numbers or occupy a new position are shown in boldface. For instance, to obtain P_{10} from P_9 we insert 4 into the first row of P_9. The 4 is inserted into the second column and bumps the 2 into the second row. The 2 is also inserted into the second column and bumps the 1 into the third row. The 1 is placed at the end of the third row. To obtain Q_{10} from Q_9 we must place 1 so that P_{10}

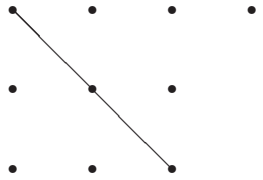
and Q_{10} have the same shape. Hence 1 is placed at the end of the third row. From the bottom entry ($i = 13$) of Figure 4 we obtain:

$$P = \begin{array}{l} 4\,4\,3\,3\,3\,1 \\ 3\,2\,2\,2\,1 \\ 1\,1 \end{array}, \qquad Q = \begin{array}{l} 3\,3\,3\,2\,2\,2 \\ 2\,2\,1\,1\,1 \\ 1\,1 \end{array}. \tag{3.4}$$

The final step of the Bender-Knuth bijection is to merge the two column-strict plane partitions P and Q into a single plane partition π. We do this by merging column-by-column, that is, the kth columns of P and Q are merged to form the kth column of π. Let us first merge the first columns of P and Q in equation (3.4). The following diagram illustrates the merging procedure:



The number of dots in each row on or to the right of the main diagonal (which runs southeast from the upper left-hand corner) is equal to 4, 3, 1, the entries of the first column of P. Similarly, the number of dots in each column on or below the main diagonal is equal to 3, 2, 1, the entries of the first column of Q. The total number of dots in each row is 4, 4, 3, and we let these numbers be the entries of the first column of π. In the same way, the second column of π has entries 4, 3, 3, as shown by the following diagram:



When this merging procedure is carried out to all the columns of P and Q, we obtain the plane partition

$$\pi = \begin{array}{l} 4\,4\,3\,3\,3\,1 \\ 4\,3\,3\,3\,2\,1 \\ 3\,3\,1 \end{array}. \tag{3.5}$$

This gives the desired bijection that proves MacMahon's formula (3.1). Of course there are many details to be proved in order to verify that this procedure has all the necessary properties. The key point is that every step is *reversible*.

i	P_i	Q_i
1	**3**	3
2	3 **1**	3 3
3	3 1 **1**	3 3 3
4	3 **2** 1 **1**	3 3 3 2
5	3 2 **2** 1 **1**	3 3 3 2 2
6	3 2 2 **2** 1 1	3 3 3 2 2 2
7	3 2 2 2 **1** 1 1	3 3 3 2 2 2 2
8	3 2 2 2 1 **1** 1 1	3 3 3 2 2 2 2 2
9	**4** 2 2 2 1 1 **3** 1 **1**	3 3 3 2 2 2 2 2 1
10	4 **4** 2 2 1 1 3 **2** 1 **1**	3 3 3 2 2 2 2 2 1 1
11	4 4 **3** 2 1 1 3 2 **2** 1 1	3 3 3 2 2 2 2 2 1 1 1
12	4 4 3 **3** 1 1 3 2 2 **2** 1 1	3 3 3 2 2 2 2 2 1 1 1 1
13	4 4 3 3 **3** 1 3 2 2 2 **1** 1 1	3 3 3 2 2 2 2 2 1 1 1 1 1

FIGURE 4

The RSK algorithm applied to the two-line array

$$A = \begin{matrix} 3 & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 3 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 4 & 4 & 3 & 3 & 3 \end{matrix}$$

The basis for this observation is that equal entries of Q are inserted left-to-right. Thus for instance in equation (3.4), the last entry to be inserted into Q is the rightmost 1, in column 5. A good way to convince yourself of the accuracy of the entire procedure is to take the plane partition π of equation (3.5) and try to reconstruct the original choice of terms from the product $1/(1-x)(1-x^2)^2\cdots$.

Extending MacMahon's theorem

By analyzing more carefully the above bijective proof, it is possible to extend the formula (3.1) of MacMahon. Write $[i]$ as short for $1-x^i$. Without going into any of the details, let us simply state that if $pp_{rs}(n)$ denotes the number of plane partitions of n with at most r rows and at most s columns, where say $r \le s$, then

$$1 + pp_{rs}(1)x + pp_{rs}(2)x^2 + \cdots =$$

$$\frac{1}{[1][2]^2[3]^3\cdots[r]^r[r+1]^r\cdots[s]^r[s+1]^{r-1}[s+2]^{r-2}\cdots[r+s-1]}. \tag{3.6}$$

For instance, when $r = 3$ and $s = 5$ the right-hand side of equation (3.6) becomes

$$\frac{1}{(1-x)(1-x^2)^2(1-x^3)^3(1-x^4)^3(1-x^5)^3(1-x^6)^2(1-x^7)}$$

$$= 1+x+3x^2+6x^3+12x^4+21x^5+39x^6+64x^7+109x^8+175x^9+280x^{10}+\cdots.$$

For example, the fact that the coefficient of x^4 is 12 means that there are 12 plane partitions of 4 with at most 3 rows and at most 5 columns. These plane partitions are given by

$$4 \quad 3\,1 \quad 2\,2 \quad 2\,1\,1 \quad 1\,1\,1\,1 \quad \begin{matrix}3\\1\end{matrix} \quad \begin{matrix}2\\2\end{matrix} \quad \begin{matrix}2\,1\\1\end{matrix} \quad \begin{matrix}1\,1\\1\,1\end{matrix} \quad \begin{matrix}1\,1\,1\\1\end{matrix} \quad \begin{matrix}2\\1\\1\end{matrix} \quad \begin{matrix}1\,1\\1\\1\end{matrix}.$$

By more sophisticated arguments (not a direct bijective proof) one can extend equation (3.6) even further, as follows. Let $\mathrm{pp}_{rst}(n)$ denote the number of plane partitions of n with at most r rows, at most s columns, and with largest part at most t. Then

$$1 + \mathrm{pp}_{rst}(1)x + \mathrm{pp}_{rst}(2)x^2 + \cdots =$$

$$\frac{[1+t][2+t]^2[3+t]^3\cdots[r+t]^r[r+1+t]^r\cdots[s+t]^r}{[1][2]^2[3]^3\cdots[r]^r[r+1]^r\cdots[s]^r}$$

$$\times\frac{[s+1+t]^{r-1}[s+2+t]^{r-2}\cdots[r+s-1+t]}{[s+1]^{r-1}[s+2]^{r-2}\cdots[r+s-1]}. \tag{3.7}$$

As a concrete example of equation (3.7), suppose that $r = 2$, $s = 3$, and $t = 2$. The right-hand side of (3.7) becomes

$$\frac{(1-x^3)(1-x^4)^2(1-x^5)^2(1-x^6)}{(1-x)(1-x^2)^2(1-x^3)^2(1-x^4)}$$

$$= 1 + x + 3x^2 + 4x^3 + 6x^4 + 6x^5 + 8x^6 + 6x^7 + 6x^8 + 4x^9 + 3x^{10} + x^{11} + x^{12}.$$

Note that the right-hand sides of equations (3.6) and (3.7) have the same denominator. The numerator of (3.7) is obtained by replacing each denominator factor $[i]$ with $[i+t]$. Equation (3.7) was also first proved by MacMahon and is the culmination of his work on plane partitions. It is closely related to some facts in *representation theory*, a subject that at first sight seems to have no connection with plane partitions. (See Chapter 5.) MacMahon's results have many other variations which give simple product formulas for enumerating various classes of plane partitions. It seems natural to try to extend these results to even higher dimensions. Thus a three-dimensional analogue of plane partitions would be *solid partitions*. All attempts (beginning in fact with MacMahon) to find nice formulas for general classes of solid partitions have resulted in failure. It seems that plane partitions are fundamentally different in behavior than their higher dimensional cousins.

The RSK algorithm has a number of remarkable properties that were not needed for the derivation of MacMahon's formula (3.1). The most striking of these properties is the following. Consider a two-line array A such as (3.3) which is the input to the RSK algorithm. Now interchange the two rows, and sort the columns so that the first row is weakly decreasing, and the part of the second row below a fixed number in the first row is also weakly decreasing. Call this new two-line array the *transposed array* A'. Thus the number of columns of A equal to $\binom{i}{j}$ is the same as the number of columns of A' equal to $\binom{j}{i}$. For the two-line array A of equation (3.3) we have

$$A' = \begin{matrix} 4 & 4 & 3 & 3 & 3 & 3 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 2 & 2 \end{matrix} \tag{3.8}$$

Thus the RSK algorithm can be applied to A'. If (P, Q) is the pair of column-strict plane partitions obtained by applying the RSK algorithm to A, then applying this algorithm to A' produces the pair (Q, P), that is, the roles of P and Q are reversed! Keeping in mind the totally different combinatorial rules for forming P and Q, it seems almost miraculous when trying a particular example such as (3.3) and (3.8) that we obtain such a simple result. We can use this "symmetry property" of the RSK algorithm to enumerate further classes of plane partitions. In particular, a plane partition is called *symmetric* if it remains the same when reflected about the main diagonal running from the upper left-hand corner in the southeast direction. An example of a symmetric plane partition is given by

$$\begin{array}{ccccccc} 5 & 3 & 3 & 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 2 & 1 & & \\ 3 & 3 & 2 & 1 & 1 & & \\ 2 & 2 & 1 & & & & \\ 1 & 1 & 1 & & & & \\ 1 & & & & & & \\ 1 & & & & & & \end{array}$$

Let $s(n)$ denote the number of symmetric plane partitions of n. For instance, $s(5) = 4$, as shown by

$$\begin{array}{ccccc} 5 & \begin{array}{c}31\\1\end{array} & \begin{array}{c}21\\11\end{array} & \begin{array}{c}111\\1\\1\end{array} & . \end{array}$$

Without going into any details, let us just say that the symmetry property of the RSK algorithm just described yields a bijective proof, similar to the proof we have given of MacMahon's formula (3.1), of the generating function

$$s(0) + s(1)x + s(2)x^2 + \cdots = \frac{1}{D(x)},$$

where

$$D(x) = (1-x)(1-x^3)(1-x^4)(1-x^5)(1-x^6)(1-x^7)(1-x^8)^2$$
$$(1-x^9)(1-x^{10})^2(1-x^{11})(1-x^{12})^3 \cdots .$$

The exponent of $1-x^{2k-1}$ in $D(x)$ is 1, and the exponent of $1-x^{2k}$ is $\lfloor k/2 \rfloor$, the greatest integer less than or equal to $k/2$.

4
Standard Young tableaux

There is a special class of objects closely related to plane partitions that are of considerable interest. Let λ be an ordinary partition of n with parts $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell$. A *standard Young tableau* (SYT) of shape λ is a left-justified array of positive integers, with λ_i integers in the ith row, satisfying the following two conditions: (1) The entries consist of the integers $1, 2, ..., n$, each occurring exactly once, and (2) the entries in each row and column are increasing. An example of an SYT of shape $(4, 3, 2)$ is given by

$$\begin{array}{llll} 1 & 3 & 4 & 6 \\ 2 & 7 & 8 & \\ 5 & 9 & & \end{array} \quad . \tag{4.1}$$

There are exactly ten SYT of size four (that is, with four entries), given by

$$\begin{array}{llllllllll} 1234 & \begin{array}{ll} 123 \\ 4 \end{array} & \begin{array}{ll} 124 \\ 3 \end{array} & \begin{array}{ll} 134 \\ 2 \end{array} & \begin{array}{ll} 12 \\ 34 \end{array} & \begin{array}{ll} 13 \\ 24 \end{array} & \begin{array}{ll} 12 \\ 3 \\ 4 \end{array} & \begin{array}{ll} 13 \\ 2 \\ 4 \end{array} & \begin{array}{ll} 14 \\ 2 \\ 3 \end{array} & \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \end{array} \end{array} .$$

THE BALLOT PROBLEM AND THE HOOK LENGTH FORMULA

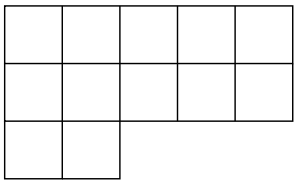
Standard Young tableaux have a number of interpretations which make them of great importance in a variety of algebraic, combinatorial, and probabilistic problems. Here we will only mention a classical problem called the *ballot*

problem, which has numerous applications in probability theory. Given a partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ as above with $\lambda_1 + \cdots + \lambda_\ell = n$, we suppose that an election is being held among ℓ candidates $A_1, \ldots, A_\ell$. At the end of the election candidate A_i receives λ_i votes. The voters vote in succession one at a time. We record the votes of the voters as a sequence $a_1, a_2, \ldots, a_n$, where $a_j = i$ if the jth voter votes for A_i. The sequence $a_1, a_2, \ldots, a_n$ is called a *ballot sequence* (of shape λ) if at no time during the voting does any candidate A_i trail another candidate A_j with $j > i$. Thus the candidates maintain their relative order (allowing ties) throughout the election. For instance, the sequence $1, 2, 1, 3, 1, 3, 4, 2$ is not a ballot sequence, since at the end A_2 and A_3 receive the same number of votes, but after six votes A_2 trails A_3. On the other hand, the sequence $1, 2, 1, 3, 1, 2, 4, 3$ is a ballot sequence. Despite the difference in their descriptions, a ballot sequence is nothing more than a disguised version of an SYT. Namely, if T is an SYT, then define $a_j = i$ if j appears in the ith row of T. A little thought should convince the reader that the sequence $a_1, a_2, \ldots, a_n$ is then a ballot sequence, and that all ballot sequences come in this way from SYT's. For instance, the SYT of equation (4.1) corresponds to the ballot sequence $1, 2, 1, 1, 3, 1, 2, 2, 3$. We are simply recording in which rows the numbers $1, 2, \ldots, 9$ appear.

It is natural (at least for a practitioner of combinatorics) to ask how many SYT there are of a given shape λ. This number is denoted f^λ. For instance, there are nine SYT of shape $(4, 2)$, which we write as $f^{4,2} = 9$. These nine SYT are given by

$$\begin{array}{ccccccccc} 1234 & 1235 & 1236 & 1245 & 1246 & 1256 & 1345 & 1346 & 1356 \\ 56 & 46 & 45 & 36 & 35 & 34 & 26 & 25 & 24 \end{array}.$$

A formula for f^λ (defined in terms of ballot sequences) was given by MacMahon in 1900. A simplified version was given by James Sutherland Frame, Robinson (mentioned earlier in connection with the RSK algorithm), and Robert McDowell Thrall in 1954, and is known as the Frame-Robinson-Thrall *hook length formula*. To state this formula, we define a *Young diagram* of shape λ as a left-justified array of squares with λ_i squares in the ith row. For instance, a Young diagram of shape $(5, 5, 2)$ looks like



An SYT of shape λ can then be regarded as an insertion of the numbers $1, 2, \ldots, n$ (each appearing once) into the squares of a Young diagram of shape λ such that every row and column is increasing. If s is a square of a Young diagram, then define the *hook length* of s to be the number of squares to the right of s and in the same row, or below s and in the same column, counting s itself once. In the following figure, we have inserted inside each square of the Young diagram of shape $(5, 5, 2)$ its hook length.

7	6	4	3	2
6	5	3	2	1
2	1			

The *hook product* H_λ of a partition λ is the product of the hook lengths of its Young diagram. Thus for instance from the above figure we see that

$$H_{5,5,2} = 7 \cdot 6 \cdot 4 \cdot 3 \cdot 2 \cdot 6 \cdot 5 \cdot 3 \cdot 2 \cdot 1 \cdot 2 \cdot 1 = 362,880.$$

The Frame-Robinson-Thrall formula can now be stated. Here λ is a partition of n and $n! = 1 \cdot 2 \cdots n$.

Hook length Formula.

$$f^\lambda = \frac{n!}{H_\lambda}. \tag{4.2}$$

For instance,

$$f^{5,5,2} = \frac{12!}{362,880} = 1320.$$

It is remarkable that such a simple formula for f^λ exists, and no really simple proof is known. The proof of Frame-Robinson-Thrall amounts to simplifying MacMahon's formula for f^λ, which MacMahon obtained by solving difference equations (the discrete analogue of differential equations). Other proofs were subsequently given, including several bijective proofs, the nicest due to the coworkers Jean-Christophe Novelli, Igor Pak and Alexander V. Stoyanovskii.

PERMUTATIONS AND STANDARD YOUNG TABLEAUX

In addition to their usefulness in combinatorics, SYT also play a significant role in a certain kind of symmetry, such as the symmetry between identical particles in quantum mechanics. This important theory (known in mathematics as "the representation theory of the symmetric group") was developed primarily by Alfred Young, who was a clergyman by profession and a fellow of Clare College, Cambridge, a Canon of Chelmsford, and Rector of Birdbrook, Essex (1910–1940). Roughly speaking, this theory describes the possible "symmetry states" of n objects. See Chapter 5 for more details.

A *permutation* of the numbers $1, 2, \ldots, n$ is simply a rearrangement, that is, a way of listing these numbers in some order. For instance, $5, 2, 7, 6, 1, 4, 3$ (also written as just 5276143 when no confusion can arise) is a permutation of $1, 2, 3, 4, 5, 6, 7$. The number of permutations of $1, 2, \ldots, n$ is $n! = n(n-1)\cdots 2 \cdot 1$. This fact was motivated in the Introduction, where we spoke about words with n distinct letters, which are easily seen to be equivalent to permutations.

It is an immediate consequence of the theory of symmetry mentioned above that the number of ordered pairs of SYT of the same shape and with n squares is equal to $n!$, i.e. the number of permutations of n objects. For instance, when $n = 3$ we get the six pairs

$$\left(\begin{array}{ccc} 1 & 2 & 3 \end{array}\right) \quad \left(\begin{array}{cc} 1 & 2 \\ 3 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 2 \\ 3 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 2 \\ 3 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 3 \\ 2 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 3 \\ 2 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 3 \\ 2 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 2 \\ 3 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 3 \\ 2 & \end{array}\right) \quad \left(\begin{array}{cc} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{array}\right).$$

The fact that the number of pairs of SYT of the same shape and with n squares is $n!$ can also be expressed by the formula

$$\sum_{\lambda \vdash n} \left(f^\lambda\right)^2 = n!, \tag{4.3}$$

where $\lambda \vdash n$ denotes that λ is a partition of n. A combinatorialist will immediately ask whether there is a bijective proof of this formula. In other words, given a permutation w of the numbers $1, 2, \ldots, n$, can we associate with w a pair (P, Q) of SYT of the same shape and with n squares, such that every such pair occurs exactly once? In fact we have already seen the solution to this problem — it is just a special case of the RSK algorithm! There is only one minor technicality that needs to be explained before we apply the RSK algorithm. Namely, the column-strict plane partitions we were dealing

$$\begin{array}{lll} 1\,3\,4 & & 9\,7\,6 \\ 2\,6\,8 & & 8\,4\,2 \\ 5\,9 & & 5\,1 \\ 7 & & 3 \end{array}$$

FIGURE 5
An SYT and its corresponding reverse SYT

with before have every row and column *decreasing*, while SYT have every row and column *increasing*. However, given a plane partition whose entries are the integers $1, 2, \ldots, n$, each appearing once (so it will automatically be column-strict), we need only replace i by $n + 1 - i$ to obtain an SYT of the same shape. We will call a plane partition whose (nonzero) parts are the integers $1, 2, \ldots, n$, each appearing once, a *reverse SYT*. An example of an SYT and the corresponding reverse SYT obtained by replacing i with $n+1-i$ is shown in Figure 5.

So consider now a permutation such as $5, 2, 6, 1, 4, 7, 3$. Write this as the second line of a two-line array whose first line is $n, n-1, \ldots, 1$. Here we get the two-line array

$$A = \begin{array}{c} 7\,6\,5\,4\,3\,2\,1 \\ 5\,2\,6\,1\,4\,7\,3 \end{array}.$$

When we apply the RSK algorithm to this two-line array, we will obtain a pair of column-strict plane partitions of the same shape whose parts are $1, 2, \ldots, n$, each appearing once. Namely, we get

$$\begin{array}{lll} 7\,4\,3 & & 7\,6\,4 \\ 6\,2\,1 & & 5\,3\,1 \\ 5 & & 2 \end{array}.$$

If we replace i by $8 - i$, we get the following pair of SYT of the same shape $(3, 3, 1)$:

$$\begin{array}{lll} 1\,4\,5 & & 1\,2\,4 \\ 2\,6\,7 & & 3\,5\,7 \\ 3 & & 6 \end{array}.$$

The process is reversible; that is, beginning with a pair (P, Q) of SYT of the same shape, we can reconstruct the permutation that produced it, as a special case of the reversibility of the RSK algorithm discussed in the previous section. Therefore the number of pairs of SYT of the same shape and with n entries is equal to the number of permutations $a_1, \ldots, a_n$ of $1, 2, \ldots, n$, yielding the

formula (4.3). This remarkable connection between permutations and tableaux is the foundation for an elaborate theory of permutation enumeration. In Chapter 6 we give a taste of this theory.

5
Connections with representation theory

In this chapter, which is independent from the rest of this book, we assume familiarity with the fundamentals of representation theory.

First we consider the group $G = \mathrm{GL}(n, \mathbf{C})$ of all invertible linear transformations on an n-dimensional complex vector space V. We will identify G with the group of $n \times n$ invertible complex matrices. A *polynomial representation* of G of degree N is a homomorphism $\varphi : G \to \mathrm{GL}(N, \mathbf{C})$, such that for $A \in G$, the entries of the matrix $\varphi(A)$ are polynomials (independent of the choice of A) in the entries of A. For instance, one can check directly that the map $\varphi : \mathrm{GL}(2, \mathbf{C}) \to \mathrm{GL}(3, \mathbf{C})$ defined by

$$\varphi \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 & 2ab & b^2 \\ ac & ad+bc & bd \\ c^2 & 2cd & d^2 \end{bmatrix} \tag{5.1}$$

preserves multiplication (and the identity element), and hence is a polynomial representation of $\mathrm{GL}(2, \mathbf{C})$ of degree 3.

Let $\varphi : \mathrm{GL}(n, \mathbf{C}) \to \mathrm{GL}(N, \mathbf{C})$ be a polynomial representation. If the eigenvalues of A are $x_1, \ldots, x_n$, then the eigenvalues of $\varphi(A)$ are *monomials* in the x_i's. For instance, in equation (5.1) one can check that if x_1 and x_2 are the eigenvalues of A, then the eigenvalues of $\varphi(A)$ are x_1^2, $x_1 x_2$, and x_2^2. The *trace* of $\varphi(A)$ (the sum of the eigenvalues) is therefore a polynomial

in the x_i's which is a sum of N monomials. This polynomial is called the *character* of φ, denoted $\mathrm{char}(\varphi)$. For φ as in (5.1), we have

$$\mathrm{char}(\varphi) = x_1^2 + x_1x_2 + x_2^2.$$

Some of the basic facts concerning the characters of $\mathrm{GL}(n, \mathbf{C})$ are the following:

- Every polynomial representation (assumed finite-dimensional) of the group $\mathrm{GL}(n, \mathbf{C})$ is completely reducible, i.e., a direct sum of irreducible polynomial representations. The multiset of these irreducible constituents is unique up to equivalence, that is, up to the choice of basis used to represent the linear transformations $\varphi(A)$, $A \in G$, as matrices.
- The characters of irreducible representations are homogeneous symmetric functions in the variables $x_1, \ldots, x_n$, and only depend on the representation up to equivalence.
- The characters of inequivalent irreducible representations are linearly independent over $\mathbf{C}$.

The effect of these properties is that once we determine the character of a polynomial representation φ of $\mathrm{GL}(n, \mathbf{C})$, then there is a unique way to write this character as a sum of irreducible characters. The representation φ is determined up to equivalence by the multiplicity of each irreducible character in $\mathrm{char}(\varphi)$. Hence we are left with the basic question of describing the irreducible characters of $\mathrm{GL}(n, \mathbf{C})$. The main result is the following.

Fundamental theorem on the polynomial characters of $\mathrm{GL}(n, \mathbf{C})$. *The irreducible polynomial characters of* $\mathrm{GL}(n, \mathbf{C})$ *are in one-to-one correspondence with the partitions* $\lambda = (\lambda_1, \ldots, \lambda_n)$ *with at most* n *parts. The irreducible character* $s_\lambda = s_\lambda(x_1, \ldots, x_n)$ *corresponding to* λ *is given by*

$$s_\lambda(x_1, \ldots, x_n) = \sum_T x^T,$$

where T *ranges over all column-strict plane partitions (as defined in Chapter 3) of shape* λ *and largest part at most* n*, and where* x^T *denotes the monomial*

$$x^T = x_1^{\text{number of 1's in } T} x_2^{\text{number of 2's in } T} \cdots .$$

For instance, let $n = 2$ and let $\lambda = (2, 0)$ be the partition with just one part equal to two (and no other parts). The column-strict plane partitions of shape $(2, 0)$ with largest part at most 2 are just 11, 21, and 22. Hence (abbreviating $s_{(2,0)}$ as s_2),

$$s_2(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2.$$

This is just the character of the representation defined by equation (5.1). Hence this representation is one of the irreducible polynomial representations of $\mathrm{GL}(2,\mathbf{C})$.

As another example, suppose that $n = 3$ and $\lambda = (2,1,0)$. The corresponding column-strict plane partitions are

$$\begin{array}{cccccccc} 21 & 22 & 31 & 31 & 32 & 32 & 33 & 33 \\ 1 & 1 & 1 & 2 & 1 & 2 & 1 & 2 \end{array}.$$

Hence

$$s_\lambda(x_1,x_2,x_3) = x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + 2x_1x_2x_3 + x_2^2x_3 + x_1x_3^2 + x_2x_3^2.$$

The fact that we have eight column-strict plane partitions in this case is closely related to the famous "Eightfold Way" of particle physics. (The corresponding representation of $\mathrm{GL}(3,\mathbf{C})$, when restricted to $\mathrm{SL}(3,\mathbf{C})$, is just the adjoint representation of $\mathrm{SL}(3,\mathbf{C})$.)

The symmetric functions $s_\lambda(x_1,\ldots,x_n)$ are known as *Schur functions* (in the variables $x_1,\ldots,x_n$) and play an important role in many aspects of representation theory, the theory of symmetric functions, and enumerative combinatorics. In particular, they are closely related to the irreducible representations of a certain *finite* group, namely, the symmetric group $\mathfrak{S}_k$ of all permutations of the set $\{1,2,\ldots,k\}$. This relationship is best understood by a "duality" between $\mathrm{GL}(n,\mathbf{C})$ and $\mathfrak{S}_k$ discovered by Issai Schur.

Recall that we are regarding $\mathrm{GL}(n,\mathbf{C})$ as acting on an n-dimensional vector space V. Thus $\mathrm{GL}(n,\mathbf{C})$ also acts on the kth tensor power $V^{\otimes k}$ of V. On the other hand, the group $\mathfrak{S}_k$ acts on $V^{\otimes k}$ by permuting tensor coordinates. Schur's famous "double centralizer" theorem asserts that the actions of $\mathrm{GL}(n,\mathbf{C})$ and $\mathfrak{S}_k$ centralize each other, i.e., every endomorphism of $V^{\otimes k}$ commuting with the action of $\mathrm{GL}(n,\mathbf{C})$ is a linear combination of the actions of the elements of $\mathfrak{S}_k$, and *vice versa*. From this one can show that the action of the group $\mathfrak{S}_k \times \mathrm{GL}(n,\mathbf{C})$ on $V^{\otimes k}$ breaks up into irreducible constituents in the form

$$V^{\otimes k} = \coprod_\lambda \left(M^\lambda \otimes F_\lambda\right), \tag{5.2}$$

where (a) $\coprod$ denotes a direct sum of vector spaces, (b) λ ranges over all partitions of k into at most n parts, (c) F_λ is the irreducible $\mathrm{GL}(n,\mathbf{C})$-module corresponding to λ, and M^λ is an irreducible $\mathfrak{S}_k$-module. Thus when $k \le n$, λ ranges over *all* partitions of k. The $p(k)$ irreducible $\mathfrak{S}_k$-modules M^λ are pairwise nonisomorphic and account for all the irreducible $\mathfrak{S}_k$-modules. Hence the irreducible $\mathfrak{S}_k$-modules are naturally indexed by partitions of k. Using the RSK algorithm (or otherwise), it is easy to prove the identity

$$(x_1 + x_2 + \cdots + x_n)^k = \sum_\lambda f^\lambda s_\lambda(x_1, \ldots, x_n),$$

where λ ranges over all partitions of k and f^λ denotes as usual the number of SYT of shape λ. Taking the $\mathrm{GL}(n, \mathbf{C})$ character of equation (5.2), the left-hand side becomes $(x_1 + \cdots + x_n)^k$, while the right-hand side becomes $\sum_\lambda (\dim M^\lambda) s_\lambda(x_1, \ldots, x_n)$. Since the Schur functions are linearly independent, it follows that $\dim M^\lambda = f^\lambda$. Thus the f^λ's for λ a partition of k are the degrees of the irreducible representations of $\mathfrak{S}_k$. Since the sum of the squares of the degrees of the irreducible representations of a finite group G is equal to the order (number of elements) of G, we obtain equation (4.3) (with n replaced by k).

We have given only the briefest glimpse of the connections between tableau combinatorics and representation theory, but we hope that it gives the reader with sufficient mathematical background some of the flavor of this subject.

6
Increasing and decreasing subsequences

In this chapter we discuss an unexpected connection between the RSK algorithm and the enumeration of a certain class of permutations. This connection was discovered by Schensted and was his reason for inventing his famous correspondence. If $w = a_1 a_2 \cdots a_n$ is a permutation of $1, 2, \ldots, n$, then a *subsequence* v of length k of w is a sequence of k distinct terms of w appearing in the order in which they appear in w. In symbols, we have $v = a_{i_1} a_{i_2} \cdots a_{i_k}$, where $i_1 < i_2 < \cdots < i_k$. For instance, some subsequences of the permutation 6251743 are 2573, 174, 6, and 6251743. A subsequence $b_1 b_2 \cdots b_k$ of w is said to be *increasing* if $b_1 < b_2 < \cdots < b_k$, and *decreasing* if $b_1 > b_2 > \cdots > b_k$. For instance, some increasing subsequences of 6251743 are 67, 257, and 3, while some decreasing subsequences are 6543, 654, 743, 61, and 3.

Longest monotone subsequences

We will be interestested in the length of the *longest* increasing and decreasing subsequences of a permutation w. Denote by is(w) the length of the longest increasing subsequence of w, and by ds(w) the length of the longest decreasing subsequence. By careful inspection one sees for instance that is(6251743) $= 3$ and ds(6251743) $= 4$. It is intuitively plausible that there should be some kind of tradeoff between the values is(w) and ds(w).

If $\mathrm{is}(w)$ is small, say equal to k, then any subsequence of w of length $k+1$ must contain a pair of decreasing elements, so there are "lots" of pairs of decreasing elements. Hence we would expect $\mathrm{ds}(w)$ to be large. An extreme case occurs when $\mathrm{is}(w) = 1$. Then there is only one choice for w, namely, $n, n-1, \ldots, 1$, and we have $\mathrm{ds}(w) = n$.

How can we quantify the feeling that that $\mathrm{is}(w)$ and $\mathrm{ds}(w)$ cannot both be small? A famous result of Pál Erdős and George Szekeres, obtained in 1935, gives an answer to this question and was one of the first results in the currently very active area of *extremal combinatorics*. For more about extremal combinatorics, see Chapter 14.

Erdős-Szekeres Theorem. *Let w be a permutation of $1, 2, \ldots, n$, and let p and q be positive integers for which $n > pq$. Then either $\mathrm{is}(w) > p$ or $\mathrm{ds}(w) > q$. Moreover, this is best possible in the sense that if $n = pq$ then we can find at least one permutation w such that $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$.*

An equivalent way to formulate the Erdős-Szekeres theorem is by the inequality

$$\mathrm{is}(w) \cdot \mathrm{ds}(w) \geq n, \tag{6.1}$$

showing clearly that $\mathrm{is}(w)$ and $\mathrm{ds}(w)$ cannot both be small. For instance, both can not be less than $\sqrt{n}$, the square root of n.

After Erdős and Szekeres proved their theorem, an extremely elegant proof was given in 1959 by Abraham Seidenberg based on a ubiquitous mathematical tool known as the *pigeonhole principle*. This principle states that if $m+1$ pigeons fly into m pigeonholes, then at least one pigeonhole contains more than one pigeon. As trivial as the pigeonhole principle may sound, it has numerous nontrivial applications. The hard part in applying the pigeonhole principle is deciding what are the pigeons and what are the pigeonholes.

We can now describe Seidenberg's proof of the Erdős-Szekeres theorem. Given a permutation $w = a_1 a_2 \cdots a_n$ of $1, 2, \ldots, n$, we define numbers $r_1, r_2, \ldots, r_n$ and $s_1, s_2, \ldots, s_n$ as follows. Let r_i be the length of the longest increasing subsequence of w that ends at a_i, and similarly let s_i be the length of the longest decreasing subsequence of w that ends at a_i. For instance, if $w = 6251743$ as above then $s_4 = 3$ since the longest decreasing subsequences ending at $a_4 = 1$ are 621 and 651, of length three. More generally, we have for $w = 6251743$ that $(r_1, \ldots, r_7) = (1, 1, 2, 1, 3, 2, 2)$ and $(s_1, \ldots, s_7) = (1, 2, 2, 3, 1, 3, 4)$.

Key fact. *The n pairs $(r_1, s_1), (r_2, s_2), \ldots, (r_n, s_n)$ are all distinct.*

To see why this fact is true, suppose i and j are numbers such that $i < j$ and $a_i < a_j$. Then we can append a_j to the end of the longest increasing

subsequence of w ending at a_i to get an increasing subsequence of greater length that ends at a_j. Hence $r_j > r_i$. Similarly, if $i < j$ and $a_i > a_j$, then we get $s_j > s_i$. Therefore we cannot have both $r_i = r_j$ *and* $s_i = s_j$, which proves the key fact.

Now suppose $n > pq$ as in the statement of the Erdős-Szekeres theorem. We therefore have n distinct pairs (r_1, s_1), $(r_2, s_2), \ldots, (r_n, s_n)$ of positive integers. If every r_i were at most p and every s_i were at most q, then there are only pq possible pairs (r_i, s_i) (since there are at most p choices for r_i and at most q choices for s_i). Hence two of these pairs would have to be equal. (This is where the pigeonhole principle comes in — we are putting the "pigeon" i into the "pigeonhole" (r_i, s_i) for $1 \leq i \leq n$. Thus there are n pigeons, where $n > pq$, and at most pq pigeonholes.) But if two pairs are equal, then we contradict the key fact above. It follows that for some i either $r_i > p$ or $s_i > q$. If $r_i > p$ then there is an increasing subsequence of w of length at least $p + 1$ ending at a_i, so $\mathrm{is}(w) > p$. Similarly, if $s_i > q$ then $\mathrm{ds}(w) > q$, completing the proof of the main part of the Erdős-Szekeres theorem.

It remains to show that the result is best possible, as explained above. In other words, given p and q, we need to exhibit at least one permutation w of $1, 2, \ldots, pq$ such that $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$. It is easy to check that the following choice of w works:

$$w = (q-1)p+1, (q-1)p+2, \ldots, qp, (q-2)p+1, (q-2)p+2, \ldots, (q-1)p, \tag{6.2}$$
$$\ldots, 2p+1, 2p+2, \ldots, 3p, p+1, p+2, \ldots, 2p, 1, 2, \ldots, p.$$

For instance, when $p = 4$ and $q = 3$ we have

$$w = 9, 10, 11, 12, 5, 6, 7, 8, 1, 2, 3, 4.$$

This completes the proof of the Erdős-Szekeres theorem.

Monotone subsequences and the RSK algorithm

Though the Erdős-Szekeres theorem is very elegant, we can ask for even more information about increasing and decreasing subsequences. For instance, rather than exhibiting a single permutation w of $1, 2, \ldots, pq$ satisfying $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$, we can ask how many such permutations there are. This much harder question can be answered by using an unexpected connection between increasing and decreasing subsequences on the one hand, and the RSK algorithm on the other.

There are two fundamental properties of the RSK algorithm that are needed for our purposes. Suppose we apply the RSK algorithm to a permutation

$w = a_1 a_2 \cdots a_n$ of $1, 2, \ldots, n$, getting two reverse SYT P and Q whose parts are $1, 2, \ldots, n$. The first property we need of the RSK algorithm is a simple description of the first row of P.

Property 1. *Suppose that the first row of P is $b_1 b_2 \cdots b_k$. Then b_i is the last (rightmost) term in w such that the longest decreasing subsequence of w ending at that term has length i.*

For instance, suppose $w = 843716925$. Then

$$P = \begin{array}{l} 9\,7\,6\,5 \\ 8\,3\,2 \\ 4\,1 \end{array}.$$

The first row of P is 9765. Consider the third element of this row, which is 6. Then 6 is the rightmost term of w for which the longest decreasing subsequence of w ending at that term has length three. Indeed, 876 is a decreasing subsequence of length three ending at 6, and there is none longer. The terms to the right of 6 are 9, 2, and 5. The longest decreasing subsequences ending at these terms have length 1, 4, and 4, respectively, so 6 is indeed the rightmost term for which the longest decreasing subsequence ending at that term has length three.

See the appendix to this chapter for a proof by induction of Property 1.

The second property we need of the RSK algorithm was first proved by Schensted. To describe this property we require the following definition. If λ is a partition, then the *conjugate* partition λ' of λ is the partition whose Young diagram is obtained by interchanging the rows and columns of the Young diagram of λ. In other words, if $\lambda = (\lambda_1, \lambda_2, \ldots)$, then the *column* lengths of the Young diagram of λ' are $\lambda_1, \lambda_2, \ldots$. For instance, if $\lambda = (5, 3, 3, 2)$ then $\lambda' = (4, 4, 3, 1, 1)$, as illustrated in Figure 6.

Property 2. Suppose that when the RSK algorithm is applied to a permutation $w = a_1 a_2 \cdots a_n$, we obtain the pair (P, Q) of reverse SYT. Let $\overline{w} = a_n a_{n-1} \cdots a_1$, the *reverse* permutation of w. Suppose that when the RSK algorithm is applied to $\overline{w}$, we obtain the pair $(\overline{P}, \overline{Q})$ of reverse SYT. Then the shape of $\overline{P}$ (or $\overline{Q}$) is conjugate to the shape of P (or Q).

Actually, an even stronger result than Property 2 is true, though we don't need it for our purposes. The reverse SYT $\overline{P}$ is actually the *transpose* of P, obtained by interchanging the rows and columns of P. (The connection between Q and $\overline{Q}$ is more subtle and has led to much interesting work.) The proof of Property 2 is too complicated for inclusion here, though it is entirely elementary.

We now have all the ingredients to state the main result (due to Schensted) on longest increasing and decreasing subsequences. If we apply the RSK

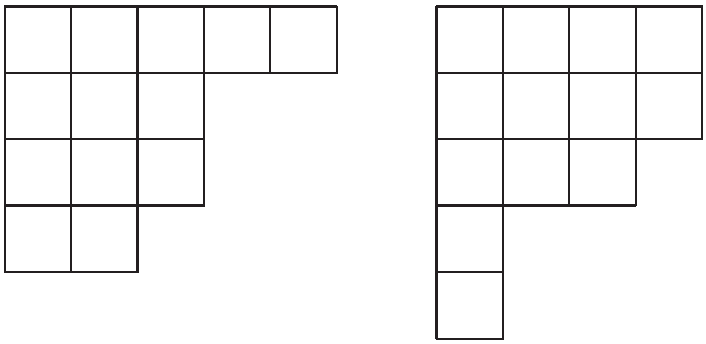


FIGURE 6
The Young diagram of a partition and its conjugate

algorithm to the permutation w and get a pair (P, Q) of reverse SYT of shape $\lambda = (\lambda_1, \lambda_2, \dots)$, then Property 1 tells us that

$$\mathrm{ds}(w) = \lambda_1.$$

In words, the length of the longest decreasing subsequence of w is equal to the largest part of λ (the length of the first row of P). Now apply the RSK algorithm to the reverse permutation $\overline{w}$, obtaining the pair $(\overline{P}, \overline{Q})$ of reverse SYT. When we reverse a permutation, increasing subsequences are changed to decreasing subsequences and *vice versa*. In particular, $\mathrm{ds}(\overline{w}) = \mathrm{is}(w)$. By Property 1, $\mathrm{ds}(\overline{w})$ is just the length of the first row of $\overline{P}$. By Property 2, the length of the first row of $\overline{P}$ is just the length of the first *column* of P. Thus $\mathrm{is}(w) = \ell(\lambda)$, the number of parts of λ.

We have shown that for a permutation w with $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$, the shape λ of the corresponding reverse SYT P (and Q) satisfies $\ell(\lambda) = p$ and $\lambda_1 = q$. Hence the number $A_n(p, q)$ of permutations w of $1, 2, \dots, n$ with $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$ is equal to the number of pairs (P, Q) of reverse SYT of the same shape λ, where λ is a partition of n with $\ell(\lambda) = p$ and $\lambda_1 = q$. How many such pairs are there? Given the partition λ, the number of choices for P is just f^λ, the number of SYT of shape λ. (Recall that the number of SYT of shape λ and the number of reverse SYT of shape λ is the same, since we can replace i by $n + 1 - i$.) Similarly there are f^λ choices for Q, so there are $\left(f^\lambda\right)^2$ choices for (P, Q). Hence we obtain our main result on increasing and decreasing subsequences:

Schensted's Theorem. *The number $A_n(p,q)$ of permutations w of $1,2,\ldots,n$ satisfying $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$ is equal to the sum of all $\left(f^\lambda\right)^2$, where λ is a partition of n satisfying $\ell(\lambda) = p$ and $\lambda_1 = q$.*

Let us see how the Erdős-Szekeres theorem follows immediately from Schensted's theorem. If a partition λ of n satisfies $\ell(\lambda) = p$ and $\lambda_1 = q$, then

$$\begin{aligned} n &= \lambda_1 + \lambda_2 + \cdots + \lambda_p \\ &\le q + q + \cdots + q \quad (p \text{ terms in all}) \\ &= pq. \end{aligned}$$

Hence if $n > pq$, then either $\ell(\lambda) \ge p+1$ or $\lambda_1 \ge q+1$. If we apply the RSK algorithm to a permutation w of $1,2,\ldots,n$ then we get a pair of reverse SYT of some shape λ, where λ is a partition of n. We have just shown that $\ell(\lambda) \ge p+1$ or $\lambda_1 \ge q+1$, so by Schensted's theorem either $\mathrm{is}(w) \ge p+1$ or $\mathrm{ds}(w) \ge q+1$.

We can evaluate each f^λ appearing in Schensted's theorem by the hook-length formula. Hence the theorem is most interesting when there are few partitions λ satisfying $\ell(\lambda) = p$ and $\lambda_1 = q$. The most interesting case occurs when $n = pq$. The fact that there is at least *one* permutation satisfying $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$ (when $n = pq$) shows that the Erdős-Szekeres theorem is best possible (see equation (6.2)). Now we are asking for a much stronger result — how many such permutations are there? By Schensted's theorem, we first need to find all partitions λ of n such that $\ell(\lambda) = p$ and $\lambda_1 = q$. Clearly there is only one such partition, namely, the partition with p parts all equal to q. Hence for this partition λ we have $A_n(p,q) = \left(f^\lambda\right)^2$. We may assume for definiteness that $p \le q$ (since $A_n(p,q) = A_n(q,p)$). In that case the hook-lengths of λ are given by 1 (once), 2 (twice), 3 (three times), $\ldots$, p (p times), $p+1$ (p times), $\ldots$, q (p times), $q+1$ ($p-1$ times), $q+2$ ($p-2$ times), $\ldots$, $p+q-1$ (once). We finally obtain the remarkable formula (for $n = pq$)

$$A_n(p,q) = \left[\frac{(pq)!}{1^1 2^2 \cdots p^p (p+1)^p \cdots q^p (q+1)^{p-1} (q+2)^{p-2} \cdots (p+q-1)^1} \right]^2.$$

For instance, when $p = 4$ and $q = 6$ we easily compute that

$$\begin{aligned} A_{24}(4,6) &= \left[\frac{24!}{1^1\, 2^2\, 3^3\, 4^4\, 5^4\, 6^4\, 7^3\, 8^2\, 9^1} \right]^2 \\ &= 19,664,397,929,878,416. \end{aligned}$$

This large number is still only a small fraction .00000003169 of the total number of permutations of $1, 2, \ldots, 24$.

Typical shape of permutations

Much more can be said about increasing and decreasing subsequences. In particular, rather than asking how many permutations w have a specified property (such as satisfying $\mathrm{is}(w) = p$ and $\mathrm{ds}(w) = q$), we can ask what is the *typical* behavior of a permutation with respect to a specified property. The study of typical behavior is a wide-ranging mathematical subject with many important applications. For instance, the whole insurance industry is based on the typical behavior of various phenomena such as heart attacks, floods, etc. We will consider here the question of the typical (or average) value of the length of the longest increasing subsequence of a permutation w of $1, 2, \ldots, n$. This question actually has applications to such topics as airplane boarding and retrieving of information on a computer.

For instance, as a very crude model of airplane boarding, suppose that a plane has seats $1, 2 \ldots, n$ from front to back. The passengers board in the order $w = a_1 a_2 \cdots a_n$ (a permutation of $1, 2, \ldots, n$). During the first unit of time, each passenger goes to his seat and sits down if possible, and otherwise waits behind other passengers. For instance, if the boarding order is 253614, then 2 and 1 go to their seats and sit down, while 536 wait behind 2 and 4 waits behind 1. The process repeats itself for the waiting passengers. Thus for the example 253614, during the second unit of time 5 and 3 go to their seats and sit down, while 64 waits behind 3. Continue until all passengers are seated. It is easy to see that the amount of time to seat all passengers is $\mathrm{is}(w)$, the length of the longest increasing subsequence of w. Eitan Bachmat and his collaborators have developed a much more sophisticated and realistic model, but the theory of increasing subsequences continues to play a central role in its analysis.

Let $E(n)$ denote the average value of $\mathrm{is}(w)$ for w a permutation of $1, 2, \ldots, n$. In other words, we obtain $E(n)$ by summing $\mathrm{is}(w)$ over all permutations w of $1, 2, \ldots, n$ and then dividing by $n!$, the total number of such permutations. For instance, when $n = 3$ we obtain the following table.

w	is(w)
123	3
132	2
213	2
231	2
312	2
321	1

Hence

$$E(3) = \frac{1}{6}(3+2+2+2+2+1) = 2.$$

How large do we expect $E(n)$ to be in general? A simple argument based on the Erdős-Szekeres theorem in the form given by equation (6.1) shows that $E(n)$ cannot be too small. Namely, given a permutation $w = a_1 a_2 \cdots a_n$ of $1, 2, \ldots, n$, let w^r denote its reverse $a_n \cdots a_2 a_1$ as was done earlier in this chapter. Since reversing a permutation converts an increasing subsequence to a decreasing one and vice versa, we have

$$\text{is}(w^r) = \text{ds}(w).$$

Hence

$$\text{is}(w) + \text{is}(w^r) = \text{is}(w) + \text{ds}(w).$$

Now for any two real numbers $x, y \geq 0$, a fundamental inquality known as the *arithmetic-geometric mean inequality* asserts that

$$\frac{x+y}{2} \geq \sqrt{xy}. \tag{6.3}$$

This inequality follows from $(x-y)^2 \geq 0$ (the square of any real number is nonnegative) by expanding out the left-hand side to get $x^2 - 2xy + y^2 \geq 0$, adding $4xy$ to both sides and dividing by 4 to get

$$\frac{x^2 + 2xy + y^2}{4} \geq xy,$$

and then taking the square root of both sides. Since $x^2 + 2xy + y^2 = (x+y)^2$, we obtain (6.3). Letting $x = \text{is}(w)$ and $y = \text{is}(w^r) = \text{ds}(w)$, we obtain

$$\frac{\text{is}(w) + \text{is}(w^r)}{2} \geq \sqrt{\text{is}(w)\text{ds}(w)}.$$

By equation (6.1) there follows

$$\frac{\text{is}(w) + \text{is}(w^r)}{2} \geq \sqrt{n}. \tag{6.4}$$

When $n > 1$ we can divide all $n!$ permutations of $1, 2, \ldots, n$ into $n!/2$ pairs w and w^r. For instance, when $n = 3$ we have the three pairs $\{123, 321\}$, $\{132, 231\}$, and $\{213, 312\}$. Equation (6.4) asserts that the average length of the longest increasing subsequence of the two permutations within each pair is at least $\sqrt{n}$. Hence the average of is(w) for *all* permutations w of $1, 2, \ldots, n$ is also at least $\sqrt{n}$. We have therefore shown that

$$E(n) \geq \sqrt{n}. \tag{6.5}$$

The question now arises as to how close $\sqrt{n}$ is to the actual value of $E(n)$. By a more sophisticated but still elementary argument, John Michael Hammersley established the upper bound

$$E(n) \leq e\sqrt{n}, \tag{6.6}$$

where $e = 2.718\cdots$ denotes the base of the natural logarithm. Equations (6.5) and (6.6) show that $E(n)$ is "about" $\sqrt{n}$, but can we do better? Is there some constant c such that $E(n)$ is close to $c\sqrt{n}$ in the sense that the ratio $E(n)/c\sqrt{n}$ gets closer and closer to 1 as n becomes larger and larger? In the language of calculus, we want $\lim_{n\to\infty} E(n)/c\sqrt{n} = 1$. Equation (6.5) shows that if c exists, then it satisfies $c \geq 1$, while equation (6.6) shows that $c \leq e$. The value of c was finally obtained independently in 1977 by Sergey Kerov and Anatoly Moiseevich Vershik, on the one hand, and Ben Logan and Lawrence Shepp on the other. Kerov and Vershik found an ingenious argument based on the RSK algorithm that $c \leq 2$, while Kerov-Vershik and Logan-Shepp showed that $c \geq 2$. It follows that $c = 2$, so $2\sqrt{n}$ is a good approximation of $E(n)$.

The argument that $c \geq 2$ is quite sophisticated, and we will mention here only one interesting aspect of it. Recall that if we apply the RSK algorithm to a permutation w of $1, 2, \ldots, n$, obtaining a pair (P, Q) of SYT of some shape λ, then is(w) is equal to λ_1, the first (or largest) part of λ. We call λ the *shape* of the permutation w. Because of this connection between increasing subsequences and RSK, it turns out that we can get information about $E(n)$ by determining what the shape λ "typically" looks like. It is not unreasonable to expect that we can make precise the idea of a "typical" shape λ obtained via the RSK algorithm, and that for this typical shape the largest part λ_1 will be near $E(n)$.

In order to talk about typical shapes of permutations of $1, 2, \ldots, n$ for varying values of n, we should scale the shapes so they have the same total size, i.e., the same total area of their boxes. If λ is partition of n, then the Young diagram of λ has n boxes. If we take the side length of each box to be $1/\sqrt{n}$, then each box has area $1/n$. Thus the total area of the boxes is 1.

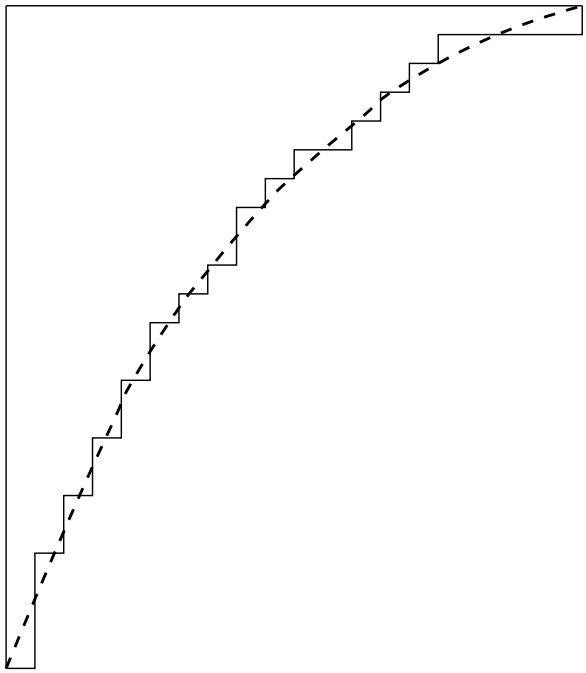


FIGURE 7
A large Young diagram

Figure 7 shows the outline of a Young diagram of a partition of $n = 154$, where we regard each box as having side length $1/\sqrt{154} = 0.08058\cdots$. The boundary of this Young diagram is approximately a curve indicated by a dashed line.

We can ask what is this curve (if it exists) for the *typical* shape of a permutation of $1, 2, \ldots, n$ when n is very large. The result of Vershik-Kerov and Logan-Shepp is that indeed such a curve Ψ exists and can be described explicitly. In other words, if a permutation w of $1, 2, \ldots, n$ is picked at random for n large, then almost certainly (in a sense that can be made precise) the outer boundary of the shape of w will be very close (again in a sense that can be made precise) to a certain curve Ψ. Thus almost all permutations of $1, 2, \ldots, n$ have approximately the same shape, a very surprising result! The curve Ψ is shown in Figure 8 (rotated 90° counterclockwise).

The curve Ψ turns out to be a new curve never previously encountered by mathematicians. The length of the line segment from the point marked 0

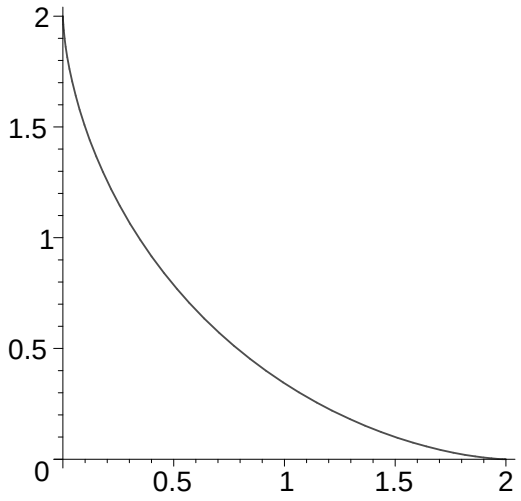


FIGURE 8
The curve $\Psi(x)$

to the intersection of the curve with the vertical line from 0 turns out to be 2. This corresponds to the length of the largest part of λ, where we have scaled each box side to have length $1/\sqrt{n}$. Hence there are typically $2\sqrt{n}$ boxes in the first row of the Young diagram of λ. This means that $\mathrm{is}(w)$ is typically about $2\sqrt{n}$, suggesting that the average $E(n)$ of $\mathrm{is}(w)$ is also close to $2\sqrt{n}$. Vershik-Kerov and Logan-Shepp made this argument completely rigorous, and thereby proved that indeed $c = 2$.

NOTE. For readers with sufficient mathematical background, we can precisely describe the curve Ψ. Namely, Ψ is given parametrically by the equations

$$\begin{aligned} x &= y + 2\cos\theta \\ y &= \frac{2}{\pi}(\sin\theta - \theta\cos\theta), \end{aligned}$$

where $0 \le \theta \le \pi$.

Although the result of Vershik-Kerov and Logan-Shepp was a big breakthrough in the theory of increasing subsequences, more exciting developments were yet to come. The number $E(n)$ merely gives the average value of $\mathrm{is}(w)$ for w a permutation of $1, 2, \ldots, n$. We can get greedy and hope to obtain even more information about $\mathrm{is}(w)$. How are the values of $\mathrm{is}(w)$ distributed

about the average value ? A definitive answer to this question was given in a seminal paper by Jinho Baik, Percy Deift, and Kurt Johansson in 1999. Their result is too sophisticated to state here, but we can explain an interesting consequence. Namely, what can be said about the error in approximating $E(n)$ by $2\sqrt{n}$? The answer is that a good approximation to the error $E(n)-2\sqrt{n}$ is $\beta n^{1/6}$, where β is a certain constant given numerically as $-1.7711\cdots$. Thus the error term is on the order of $n^{1/6}$ (the sixth root of n). More precisely, in the language of calculus we have

$$\lim_{n\to\infty} \frac{E(n)-2\sqrt{n}}{n^{1/6}} = \beta.$$

As n gets larger and larger, the ratio $(E(n)-2\sqrt{n})/n^{1/6}$ gets closer and closer to β.

For the benefit of readers familiar with calculus, we will give the definition of the constant β. The subtlety of this definition is an indication of the depth of the result of Baik, Deift, and Johansson. Let $u(x)$ denote the unique solution to the nonlinear second order equation

$$u''(x) = 2u(x)^3 + xu(x),$$

subject to the condition

$$u(x) \sim -\mathrm{Ai}(x), \text{ as } x \to \infty.$$

Here $\mathrm{Ai}(x)$ denotes the Airy function, a well-known "higher transcendental function". Define for all real numbers t the function (known as the *Tracy-Widom distribution*)

$$F(t) = \exp\left(-\int_t^\infty (x-t)u(x)^2\,dx\right).$$

Then

$$\beta = \int_{-\infty}^{\infty} t\,dF(t).$$

APPENDIX

Proof of Property 1. Recall that $w = a_1 a_2 \cdots a_n$. We prove by induction on j that after the RSK algorithm has been applied to $a_1 a_2 \cdots a_j$, yielding a pair (P_j, Q_j) of column-strict plane partitions, then the ith entry in the first row of P_j is the rightmost term of the sequence $a_1 a_2 \cdots a_j$ such that the longest decreasing subsequence ending at that term has length i. Once this is proved, then set $j = n$ to obtain Property 1.

The assertion is clearly true for $j = 1$. Assume it true for j. Suppose that the first row of P_j is $c_1 c_2 \cdots c_r$. By the induction hypothesis, c_i is the rightmost term of the sequence $a_1 a_2 \cdots a_j$ such that the longest decreasing subsequence ending at that term has length i. We now insert a_{j+1} into the first row of P_j according to the rules of the RSK algorithm. It will bump the leftmost element c_i of this row which is less than a_{j+1}. (If there is no element of the first row of P_j which is less than a_{j+1}, then a_{j+1} is inserted at the end of the row. We then set $i = r+1$, so that a_{j+1} is in all cases the ith element of the first row of P_{j+1}.) We need to show that the longest decreasing subsequence of the sequence $a_1 a_2 \cdots a_{j+1}$ ending at a_{j+1} has length i, since clearly a_{j+1} will be the *rightmost* element of $a_1 a_2 \cdots a_{j+1}$ with this property (since it is the rightmost element of the entire sequence).

If $i = 1$, then a_{j+1} is the largest element of the sequence $a_1 a_2 \cdots a_{j+1}$, so the longest decreasing subsequence ending at a_{j+1} has length one, as desired. If $i > 1$, then there is a decreasing subsequence of $a_1 a_2 \cdots a_j$ of length $i-1$ ending at c_{i-1}. Adjoining a_{j+1} to the end of this subsequence produces a decreasing subsequence of length i ending at a_{j+1}. It remains to show that there cannot be a longer decreasing subsequence ending at a_{j+1}. If there were, then there would be some term a_s in w to the left of a_{j+1} and larger than a_{j+1} such that the longest decreasing subsequence ending at a_s has length i. Thus when a_s is inserted into P_{s-1} during the RSK algorithm, it becomes the ith element of the first row. It can only be bumped by terms *larger* than a_s. In particular, when a_{j+1} is inserted into the first row, the ith element is larger than a_s, which is larger than a_{j+1}. This contradicts the definition of the bumping procedure and completes the proof.

7
Reduced decompositions

In this chapter we explore a remarkable and unexpected connection between standard Young tableaux and the building up of a permutation by interchanging (transposing) two adjacent entries.

REDUCED DECOMPOSITIONS

We begin with the *identity permutation* $1, 2, \ldots, n$. We wish to construct from it a given permutation as quickly as possible by interchanging adjacent elements. By "as quickly as possible," we mean in as few interchanges (called *adjacent transpositions*) as possible. This will be the case if we always transpose two elements a, b appearing in ascending order. For instance, one way to get the permutation 41352 from 12345 with a minimum number of adjacent transpositions is as follows, where we have marked in boldface the pair of elements to be interchanged:

$$1\mathbf{23}45 \to 13\mathbf{24}5 \to 1\mathbf{34}25 \to \mathbf{14}325 \to 413\mathbf{25} \to 41352. \tag{7.1}$$

Such sequences of interchanges are used in some of the *sorting algorithms* studied in computer science (see Chapter 15), although there it is natural to consider the reverse process whereby a list of numbers such as 41352 is step-by-step converted to the "sorted" list 12345. Note that the five steps in the sequence (7.1) are the minimum possible, since in the final permutation

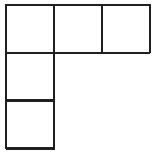
41352 there are five pairs (i,j) out of order, i.e., i appears to the left of j and $i > j$ (namely, $(4,1),(4,3),(4,2),(3,2),(5,2)$), and each adjacent transposition can make at most one pair which was in order go out of order. It would be inefficient to transpose a pair (a,b) that is in order in the final permutation, since we would only have to change it back later. A pair of elements of a permutation w that is out of order is called an *inversion* of w. The number of inversions of w is denoted $\mathrm{inv}(w)$ and is an important invariant of a permutation, in a sense measuring how "mixed up" the permutation is. For instance, $\mathrm{inv}(41352) = 5$, the inversions being the five pairs $(4,1),(4,3),(4,2),(3,2),(5,2)$.

A sequence of adjacent transpositions that converts the identity permutation to a permutation w in the smallest possible number of steps (namely, $\mathrm{inv}(w)$ steps) is called a *reduced decomposition* of w. Equation (7.1) shows one reduced decomposition of the permutation $w = 41352$, but there are many others. We can therefore ask for the number of reduced decompositions of w. We denote this number by $r(w)$. The reader can check that every permutation of the numbers $1,2,3$ has only one reduced decomposition, except that $r(321) = 2$. The two reduced decompositions of 321 are $123 \to 213 \to 231 \to 321$ and $123 \to 132 \to 312 \to 321$.

The remarkable connection between $r(w)$ and SYT's is the following. For each permutation w, one can associate a *small* collection $Y(w)$ of Young diagrams (with repetitions allowed) whose number of squares is $\mathrm{inv}(w)$, such that $r(w)$ is the sum of the number of SYT whose shapes belong to $Y(w)$. We are unable to explain here the exact rule for computing $Y(w)$, but we will discuss the most interesting special case. We also will not explain exactly what is meant by a "small" collection, but in general its number of elements will be much smaller than $r(w)$ itself.

Example. Here are a few examples of the collection $Y(w)$.

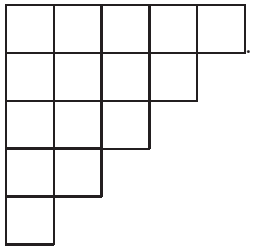
(a) If $w = 41352$ (the example considered in equation (7.1)), then $Y(w)$ consists of the single diagram



of shape $(3,1,1)$. Since there are six SYT of this shape (computed from the hook-length formula (4.2) or by direct enumeration), it follows that there are six reduced decompositions of 41352. These reduced decompositions are

$$12345 \to 12435 \to 14235 \to 41235 \to 41325 \to 41352$$
$$12345 \to 12435 \to 14235 \to 14325 \to 41325 \to 41352$$
$$12345 \to 12435 \to 14235 \to 14325 \to 14352 \to 41352$$
$$12345 \to 13245 \to 13425 \to 14325 \to 14352 \to 41352$$
$$12345 \to 13245 \to 13425 \to 14325 \to 41325 \to 41352$$
$$12345 \to 13245 \to 13425 \to 14325 \to 14352 \to 41352.$$

(b) If $w = 654321$ then again $Y(w)$ is given by a single diagram, this time



Hence

$$\begin{aligned} r(w) &= f^{(5,4,3,2,1)} \\ &= \frac{15!}{1^5 \cdot 3^4 \cdot 5^3 \cdot 7^2 \cdot 9} \\ &= 292,864. \end{aligned}$$

(c) If $w = 321654$, then $Y(w)$ consists of the diagrams whose shapes are (writing for instance 42 as short for $(4,2)$) 42, 411, 33, 321, 321, 3111, 222, 2211. Note that the shape 321 appears twice. We get

$$\begin{aligned} r(w) &= f^{42} + f^{411} + f^{33} + 2f^{321} + f^{3111} + f^{222} + f^{2211} \\ &= 9 + 10 + 5 + 2 \cdot 16 + 10 + 5 + 9 \\ &= 80. \end{aligned}$$

Clearly the formula for $r(w)$ will be the simplest when $Y(w)$ consists of a single partition λ, for then we have $r(w) = f^\lambda$, given explicitly by (4.2). A simple though surprising characterization of all permutations for which $Y(w)$ consists of a single partition is given by the next result. Such permutations are called *vexillary* after the Latin word *vexillum* for "flag," because of a relationship between vexillary permutations and certain polynomials known as *flag Schur functions*.

Vexillary theorem. *Let* $w = w_1 w_2 \cdots w_n$ *be a permutation of* $1, 2, \ldots, n$. *Then* $Y(w)$ *consists of a single partition* λ *if and only if there do not exist* $a < b < c < d$ *such that* $w_b < w_a < w_d < w_c$. *Moreover, if* α_i *is the number of* j*'s in a vexillary permutation* w *for which* $i < j$ *and* $w_i > w_j$, *then the parts of* λ *are just the nonzero* α_i*'s.*

As an illustration of the above theorem, let $w = 526314$. One sees by inspection that w satisfies the conditions of the theorem. We have $(\alpha_1, \ldots, \alpha_6) = (4, 1, 3, 1, 0, 0)$. Hence $\lambda = (4, 3, 1, 1)$ and $r(w) = f^{(4,3,1,1)} = 216$.

It is immediate from the above result that all the permutations of $1, \ldots, n$ for $n \leq 3$ are vexillary, and that there is just one nonvexillary permutation of $1, 2, 3, 4$, namely, 2143. It has been computed that if $v(n)$ denotes the number of vexillary permutations of $1, 2, \ldots, n$ then $v(5) = 103$ (out of 120 permutations of $1, 2, 3, 4, 5$ in all), $v(6) = 513$ (out of 720), $v(7) = 2761$ (out of 5040), and $v(8) = 15767$ (out of 40320). Simple methods for computing and approximating $v(n)$ have been given by Julian West and Amitai Regev, and an explicit formula for $v(n)$ was found by Ira Gessel. In particular, a good approximation to $v(n)$ (for n large) is given by

$$v(n) \sim \frac{3^{9/2} 9^n}{16\pi n^4} = 2.791102533 \cdots \frac{9^n}{n^4}.$$

Reduced decompositions and SYT of staircase shape

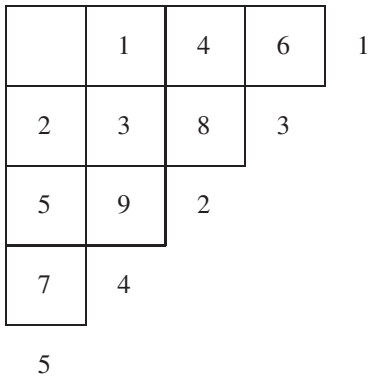
There is one class of vexillary permutations of particular interest. These are the permutations $w_0 = n, n-1, \ldots, 1$, for which $\lambda = (n-1, n-2, \ldots, 1)$. There is an elegant bijection between the SYT of shape $(n-1, n-2, \ldots, 1)$ and the reduced decompositions of w_0, due to Paul Henry Edelman and Curtis Greene. Begin with an SYT of shape $(n-1, n-2, \ldots, 1)$ and write the number i at the end of the ith row, with n written at the bottom of the first column. We will call the numbers outside the diagram *exit numbers*. An example is given by:

1	3	4	6	1
2	8	10	2	
5	9	3		
7	4			
5				

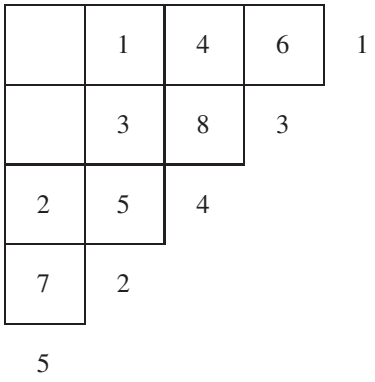
Now take the largest number in the SYT (in this case 10) and let it "exit" the diagram to the southeast (between the 2 and 3). Whenever a number exits the diagram, transpose the two exit numbers that it goes between. Hence we now have:

1	3	4	6	1
2	8		3	
5	9	2		
7	4			
5				

In the hole left by the 10, move the largest of the numbers directly to the left or above the hole. Here we move the 8 into the hole, creating a new hole. Continue to move the largest number directly to the left or above a hole into the hole, until such moves are no longer possible. Thus after exiting the 10, we move the 8, 3, and 1 successively into holes, yielding:



Now repeat this procedure, first exiting the largest number in the diagram (ignoring the exit numbers), then transposing the two exit numbers between which this largest number exits, and then filling in the holes by the same method as before. Hence for our example 9 exits, 5 fills in the hole left by 9, and 2 fills in the hole left by 5, yielding:



Continue in this manner until all the numbers are removed from the original SYT. The remarkable fact is that the exit numbers, read from top to bottom, will now be $n, n-1, \ldots, 1$. We began with the exit numbers in the order $1, 2, \ldots, n$, and each exit from the diagram transposed two adjacent exit numbers. The size (number of entries) of the original SYT is equal to $n(n-1)/2$, which is the number of inversions of the permutation $n, n-1, \ldots, 1$. Hence we have converted $1, 2, \ldots, n$ to $n, n-1, \ldots, 1$ by $n(n-1)/2$ adjacent transpositions, thereby defining a reduced decomposition of w_0. Edelman

and Greene prove that this algorithm yields a bijection between SYT of shape $(n-1, n-2, \ldots, 1)$ and reduced decompositions of w_0. For the above example, the reduced decomposition is given by $12345 \to 13245 \to 13425 \to 14325 \to 14352 \to 41352 \to 41532 \to 45132 \to 45312 \to 45321 \to 54321$.

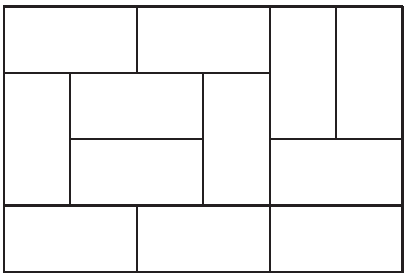
The proof of Edelman and Greene that the algorithm just defined is a bijection is quite sophisticated, and we can say only a few words about it here. Suppose that the algorithm produces the reduced decomposition R from the SYT T of shape $(n-1, n-2, \ldots, 1)$. In order to prove the bijectivity of the algorithm, it is necessary to describe how to obtain T from R. This "inverse bijection" is accomplished with a variation of the RSK algorithm. It is surprising how the RSK algorithm enters into this problem and illustrates the "robustness" of the RSK algorithm, that is, how it can be adapted to other situations. There are now known dozens of variants of the RSK algorithm, perhaps the most mathematically interesting of all combinatorial algorithms.

8
Tilings

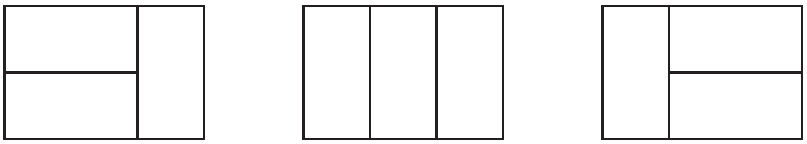
The next enumerative topic we will discuss concerns the partitioning of some planar or solid shape into smaller shapes. Such partitions are called *tilings*. The combinatorial theory of tilings is connected with such subjects as geometry, group theory, and logic, and has applications to statistical mechanics, coding theory, and many other topics. Here we will be concerned with the purely enumerative question of counting the number of tilings.

DOMINO TILINGS

The first significant result about the enumeration of tilings was due to the Dutch physicist Pieter Willem Kasteleyn and independently to the British physicist Harold Neville Vazeille Temperley and the British-born physicist Michael Ellis Fisher. Motivated by work related to the adsorption of diatomic molecules on a surface and other physical problems, they were in the early 1960s led to consider the tiling of a chessboard by dominos (or dimers). More precisely, consider an $m \times n$ chessboard B, where at least one of m and n is even. A *domino* consists of two adjacent squares (where "adjacent" means having an edge in common). The domino can be oriented either horizontally or vertically. Thus a tiling of B by dominos will require exactly $mn/2$ dominos, since there are mn squares in all, and each domino has two squares. The illustration below shows a domino tiling of a 4×6 rectangle.



Let $N(m,n)$ denote the number of domino coverings of an $m\times n$ chessboard. For instance, $N(2,3)=3$, as shown by:



We have in fact that

$$N(2,n)=F_{n+1}, \tag{8.1}$$

where F_{n+1} denotes a Fibonacci number, defined by the recurrence

$$F_1=1,\quad F_2=1,\quad F_{n+1}=F_n+F_{n-1} \text{ if } n\geq 2.$$

To prove equation (8.1), we need to show that $N(2,1)=1$, $N(2,2)=2$, and $N(2,n+2)=N(2,n+1)+N(2,n)$. Of course it is trivial to check that $N(2,1)=1$ and $N(2,2)=2$. In any domino tiling of a $2\times(n+2)$ rectangle, either the first column consists of a vertical domino, or else the first two columns consist of two horizontal dominos. In the former case we are left with a $2\times(n+1)$ rectangle to tile by dominos, and in the latter case a $2\times n$ rectangle. There are $N(2,n+1)$ ways to tile the $2\times(n+1)$ rectangle and $N(2,n)$ ways to tile the $2\times n$ rectangle, so the recurrence $N(2,n+2)=N(2,n+1)+N(2,n)$ follows, and hence also (8.1).

The situation becomes much more complicated when dealing with larger rectangles, and rather sophisticated techniques such as the "transfer-matrix method" or the "Pfaffian method" are needed to produce an answer. The final form of the answer involves trigonometric functions (see the Note below), and it is not even readily apparent (without sufficient mathematical background) that the formula gives an integer. It follows, however, from the subject known as *Galois theory* that $N(2n,2n)$ is in fact the square or twice the square of an integer, depending on whether n is even or odd. For instance,

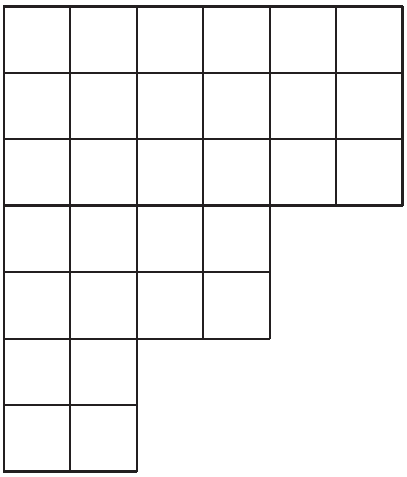


FIGURE 9
The board R_4.

$N(8,8) = 12,988,816 = 3604^2$, while $N(6,6) = 6728 = 2 \cdot 58^2$. It is natural to ask for a *combinatorial* reason why these numbers are squares or twice squares. In other words, in the case when n is even we would like a combinatorial interpretation of the number $M(2n)$ defined by $N(2n,2n) = M(2n)^2$, and similarly when n is odd. While a formula for $M(2n)$ was known making it obvious that it was an integer (so not involving trigonometric functions), it was only in 1992 that William Carl Jockusch found a direct combinatorial interpretation of $M(2n)$. In 1996 Mihai Adrian Ciucu found an even simpler interpretation of $M(2n)$ as the number of domino tilings of a certain region R_n, up to a power of two. The region R_n is defined to be the board consisting of $2n-2$ squares in the first three rows, then $2n-4$ squares in the next two rows, then $2n-6$ squares in the next two rows, etc., down to two squares in the last two rows. All the rows are left-justified. The board R_4 is illustrated in Figure 9.

If $T(n)$ denotes the number of domino tilings of R_n, then Ciucu's formula states that

$$N(2n,2n) = 2^n T(n)^2.$$

If n is even, say $n = 2r$, then $N(2n,2n) = (2^r T(n))^2$, while if n is odd, say $n = 2r+1$, then $N(2n,2n) = 2(2^r T(n))^2$, so we recover the result that

$N(2n, 2n)$ is a square or twice a square depending on whether n is even or odd.

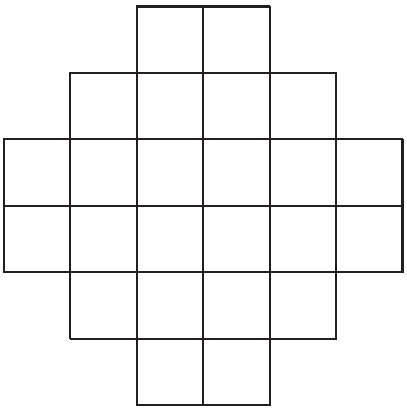
NOTE. For those readers with sufficient background, the formula for the number $N(2m, 2n)$ of domino tilings of a $2m \times 2n$ chessboard is given by

$$N(2m, 2n) = 4^{mn} \prod_{s=1}^{m} \prod_{t=1}^{n} \left(\cos^2 \frac{s\pi}{2m+1} + \cos^2 \frac{t\pi}{2n+1} \right).$$

AZTEC DIAMONDS

Although the formula for the number of domino tilings of a chessboard is rather complicated, there is a variant of the chessboard for which a very simple formula for the number of domino tilings exists. This new board is called an *Aztec diamond* and was introduced by Noam David Elkies, Gregory John Kuperberg, Michael Jeffrey Larsen, and James Gary Propp. Their work has stimulated a flurry of activity on exact and approximate enumeration of domino tilings, as well as related questions such as the appearance of a "typical" domino tiling of a given region.

The Aztec diamond AZ_n of order n consists of two squares in the first row, four squares in the second row beginning one square to the left of the first row, six squares in the third row beginning one square to the left of the second row, etc., up to $2n$ squares in the nth row. Then reflect the diagram created so far about the bottom edge and adjoin this reflected diagram to the original. For instance, the Aztec diamond AZ_3 looks as follows:



Let $\mathrm{az}(n)$ be the number of domino tilings of the Aztec diamond AZ_n. For instance, AZ_1 is just a 2×2 square, which has two domino tilings (both

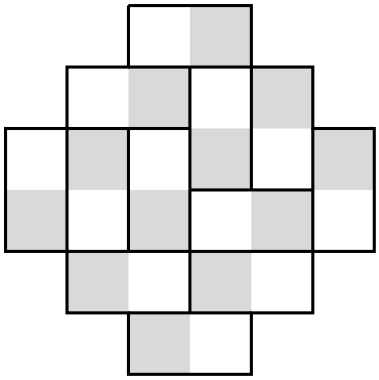


FIGURE 10
Chessboard coloring of Aztec diamond

dominos horizontal or both vertical). Hence $\mathrm{az}(1) = 2$. It's easy to compute by hand that $\mathrm{az}(2) = 8$, and a computer reveals that $\mathrm{az}(3) = 64 = 2^6$, $\mathrm{az}(4) = 1024 = 2^{10}$, $\mathrm{az}(5) = 32768 = 2^{15}$, etc. The evidence quickly becomes overwhelming for the conjecture that

$$\mathrm{az}(n) = 2^{\frac{1}{2}n(n+1)}. \tag{8.2}$$

It is rather mysterious why Aztec diamonds seem to be so much more nicely behaved regarding their number of domino tilings than the more natural $m \times n$ chessboards.

A proof of the conjecture (8.2) is the main result of Elkies *et al.* mentioned above. They gave four different proofs, showing the surprising connections between Aztec diamonds and various other branches of mathematics. (For instance, it is not a coincidence that $2^{\frac{1}{2}n(n+1)}$ is the degree of an irreducible representation of the group $\mathrm{GL}(n+1, \mathbf{C})$.) Of course a combinatorialist would like to see a purely combinatorial proof, and indeed Elkies *et al.* gave such proofs. Other combinatorial proofs have been since given by Ciucu and Propp. We will sketch the fourth proof of Elkies *et al.*, called a proof by *domino shuffling*. The domino shuffling procedure we describe will seem rather miraculous, and there are many details to verify to see that it actually works as claimed. Nevertheless, we hope that our brief description will take some of the mystery out of equation (8.2).

We first color the squares of the Aztec diamond AZ_n black and white in the usual chessboard fashion, with the first (leftmost) square in the top row

colored *white*. Figure 10 shows a tiling of AZ_3 with the chessboard coloring shown.

Certain pairs of dominos in the tiling will form a 2×2 square with the top left square colored black. Remove all such pairs of dominos (if any exist). For the tiling of AZ_3 shown above there is one such pair, and after removing it we get the tiling shown in Figure 11

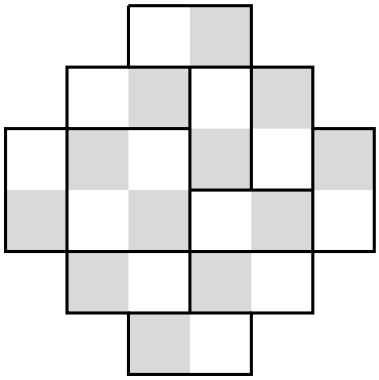


FIGURE 11
A reduced tiling

Let us call a tiling T of AZ_n with the 2×2 squares removed as just described a *reduced tiling* of AZ_n, and call T the *reduction* of the original (complete) tiling. Note that if we remove k 2×2 squares from a complete tiling to get a reduced tiling, then there are 2^k ways to tile the 2×2 holes. (Each hole can be tiled either by two horizontal or two vertical dominos.) In other words, given a reduced tiling T of AZ_n with k 2×2 holes, there are 2^k corresponding complete tilings of AZ_n whose reduction is T.

Consider a reduced tiling of AZ_n. Each domino will have one white square and one black square. There are four possible colorings and orientations of a domino, shown in Figure 12. With each of these four possible colored dominos we associate a direction: up, down, right, and left, as indicated in Figure 12 by an arrow.

We can enlarge the Aztec diamond AZ_n to AZ_{n+1} by adding squares around the boundary. Add one square at the beginning and one square at the end of each row, and two squares at the top and bottom. The next illustration shows the earlier reduced tiling of AZ_3, with an arrow placed on each domino according to its coloring and orientation, and the boundary of new squares

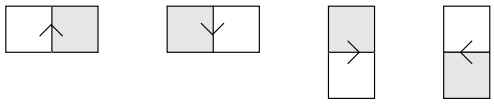


FIGURE 12
Directions of colored dominoes

to give AZ_4. We have also numbered each domino for compatibility with Figure 14.

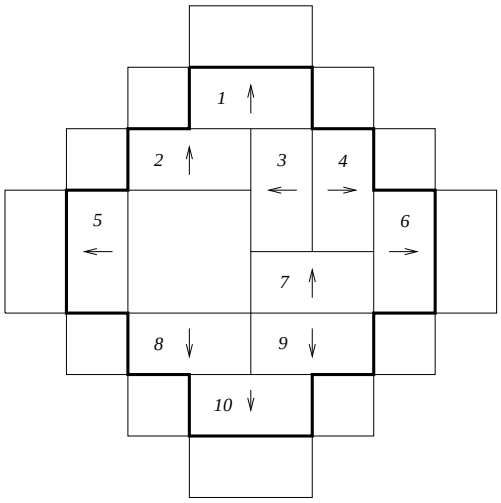


FIGURE 13
Domino shuffling

Now move each domino one unit in the direction of its arrow. This is the *shuffling* operation referred to in the name "domino shuffling." Let k be the number of 2×2 squares removed before shuffling. It can be shown that (a) the dominos do not overlap after shuffling, and (b) the squares of AZ_{n+1} that are not covered by dominos can be uniquely covered with exactly $n + k + 1$ 2×2 squares. Figure 14 shows the dominos after shuffling (with the same numbers as before), together with the leftover five 2×2 squares (holes).

We now complete the partial tiling of AZ_{n+1} to a complete tiling by putting two dominos in each 2×2 hole. Since there are two ways to tile a 2×2 square, there are 2^{n+k+1} ways to tile all $n + k + 1$ of the 2×2 squares. Therefore we have associated 2^{n+k+1} tilings of AZ_{n+1} with each

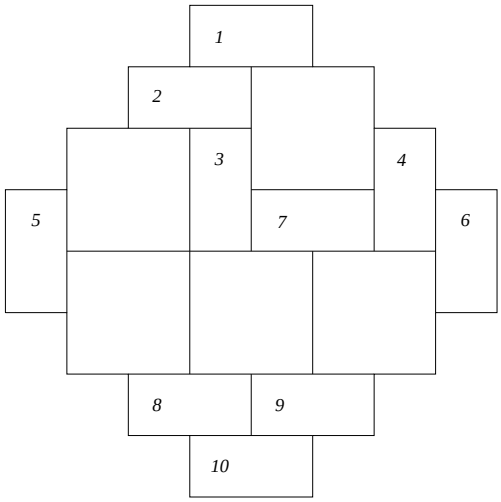


FIGURE 14
After the shuffle

k-hole reduced tiling of AZ_n. The amazing fact is that every tiling of AZ_{n+1} occurs exactly once in this way! In other words, given a tiling of AZ_{n+1}, we can reconstruct which of the dominos were shuffled from a reduced tiling of AZ_n and thus also the $n+k+1$ 2×2 holes that were left over. Since every k-hole reduced tiling T of AZ_n is the reduction of 2^k complete tilings of AZ_n, and since T corresponds to 2^{n+k+1} tilings of AZ_{n+1}, we obtain the recurrence

$$\mathrm{az}(n+1) = 2^{n+1}\mathrm{az}(n).$$

The unique solution to this recurrence satisfying $\mathrm{az}(1) = 2$ is easily seen (for instance by mathematical induction) to be

$$\mathrm{az}(n) = 2^{\frac{1}{2}n(n+1)},$$

proving equation (8.2).

We have succeeded in counting the number of domino tilings of the Aztec diamond of order n. We can go off in a different direction and ask what the tilings themselves look like. In other words, what are the properties of a "typical" tiling of AZ_n? This question is in the same spirit as asking for the typical shape of a permutation under the RSK algorithm, as we did in Chapter 6. It is by no means apparent that anything interesting can be said about a typical tiling of AZ_n. Certainly it has certain statistical properties

such as the number of horizontal dominos being approximately equal to the number of vertical dominos, but this is not very surprising and has little to do with the shape of an Aztec diamond. If, however, one looks at a tiling of a large Aztec diamond chosen randomly by a computer, some remarkable behaviour becomes evident. Figure 15 shows such a tiling of AZ_{15}, with the horizontal dominos shaded.

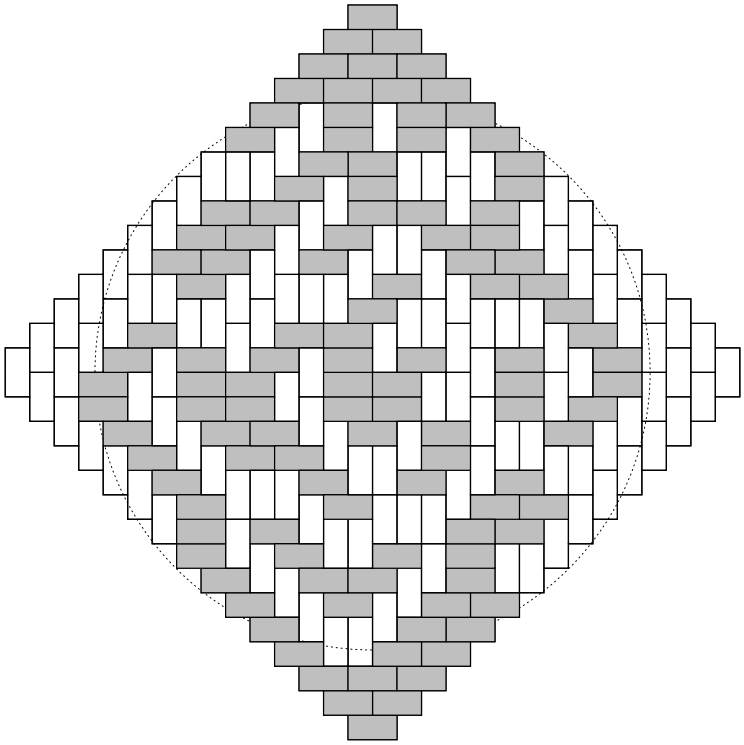


FIGURE 15
A random tiling

Note that near the four corners the dominos line up in a single direction (horizontal at the top and bottom, and vertical on the left and right), but in the middle their orientation looks random. The "chaotic region" where the dominos look random is roughly a circle. Moreover, if n is very large and the Aztec diamond is regarded as having area 1, then the boundary of the Aztec diamond will look like a square standing on one of its corners. (The actual

staircase boundary will have steps so small that they cannot be seen, so the staircases will look like straight lines.) The circle enclosing the chaotic region then appears to be tangent to the four sides of the square boundary. In 1995 it was proved by William Jockusch, James Propp, and Peter Shor that this circle indeed exists. More precisely, for very large n, almost every (in a sense that can be made precise) domino tiling of AZ_n will have dominos lined up near the four corners (horizontal at the top and bottom, and vertical on the left and right) and a chaotic region where at a particular point a domino is just as likely (to a very close approximation) to be horizontal as vertical. Moreover, this chaotic region is very close to being the interior of the circle tangent to the four (smoothed out) sides of the Aztec diamond. The circle surrounding the chaotic region is called the *Arctic circle* because the orientations of the dominos outside the circle are "frozen." This result of Jockusch et al. is a beautiful example of "typical behavior" and has inspired a lot of subsequent research.

We cannot go into the details of the proof here, but we can give a little intuition about the existence of a frozen region. Why, for instance, do dominos "want" to line up horizontally at the top? Suppose that the dominos were vertical in the top row. This forces the orientation of many other dominos, as illustrated in Figure 16.

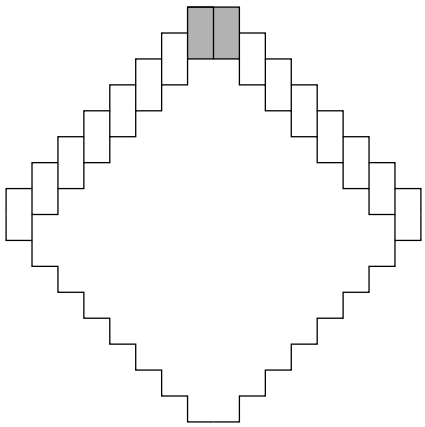


FIGURE 16
Frozen dominos

On the other hand, if the top row contains a horizontal domino then there are no such restrictions on the orientations of other dominos. Thus with the

horizontal orientation there is a considerably larger board remaining in which to place the other dominos, so we would expect many more tilings with the top row occupied by a horizontal domino. Hence a random tiling should "almost always" have a horizontal domino in the top row. Similar reasoning then applies to the next row from the top, etc., though it is by no means apparent from this heuristic reasoning where will lie the exact cutoff between "frozen" and random behavior.

9
Tilings and plane partitions

We have encountered several examples of unexpected connections between seemingly unrelated mathematical problems. This is one of the features of mathematics that makes it so appealing to its practitioners. In this chapter we discuss another such connection, this time between tilings and plane partitions. Other surprising connections will be treated in later chapters.

The tiling problem we will be considering is very similar to the problem of tiling an $m \times n$ chessboard with dominos. Instead of a chessboard (whose shape is a rectangle), we will be tiling a hexagon. Replacing the squares of the chessboard will be equilateral triangles of unit length which fill up the hexagon, yielding a "hexagonal board." Let $H(r,s,t)$ denote the hexagonal board whose opposite sides are parallel and whose side lengths (in clockwise order) are r,s,t,r,s,t. Thus opposite sides of the hexagon have equal length just like opposite sides of a rectangle have equal length. Figure 17 shows the hexagonal board $H(2,3,3)$ with its 42 equilateral triangles. In general, the hexagonal board $H(r,s,t)$ has $2(rs+rt+st)$ equilateral triangles.

Instead of tiling with dominos (which consist of two adjacent squares), we will be tiling with pieces which consist of two adjacent equilateral triangles. We will call these pieces simply *rhombi*, although they are really only special kinds of rhombi. Thus the number of rhombi in a tiling of $H(r,s,t)$ is $rs+rt+st$. The rhombi can have three possible orientations (compared with

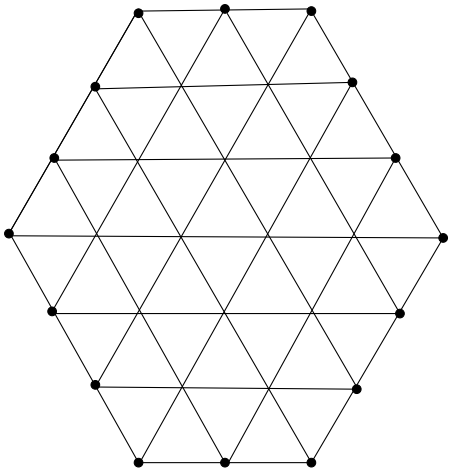
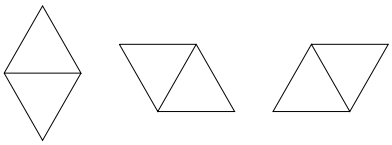
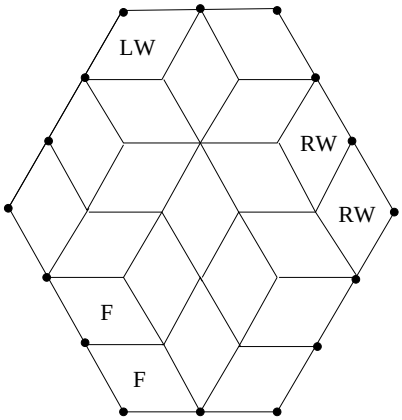


FIGURE 17
The hexagonal board $H(2,3,3)$

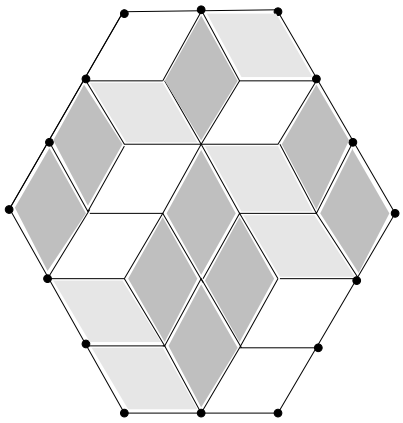
the two orientations of a rectangle):



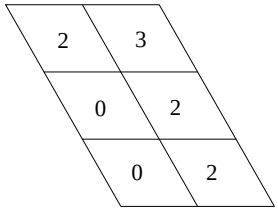
Here is a typical tiling of $H(2,3,3)$:



This picture gives the impression of looking into the corner of an $r \times s \times t$ box in which cubes are stacked. The brain will alternate between different interpretations of this cube stacking. To be definite, we have labelled by F the floor, by LW the left wall, and by RW the right wall. Shading the rhombi according to their orientation heightens the impression of a cube stacking, particularly if the page is rotated slightly counterclockwise:



Regarding the floor as a 3×2 parallelogram filled with six rhombi, we can encode the cube stacking by a 3×2 array of numbers which tell the number of cubes stacked above each floor rhombus:



Rotate this diagram 45° counterclockwise, erase the rhombi, and "straighten out," giving the following array of numbers:

$$\begin{matrix} 3\,2\,2 \\ 2\,0\,0 \end{matrix}.$$

This array is nothing more than a plane partition whose number of rows is at most r, whose number of columns is at most s, and whose largest part is at most t (where we began with the hexagonal board $H(r,s,t)$)! This correspondence between rhombic tilings of $H(r,s,t)$ and plane partitions with

at most r rows, at most s columns, and with largest part at most t is a bijection. In other words, given the rhombic tiling, there is a unique way to interpret it as a stacking of cubes (once we agree on what is the floor, left wall, and right wall), which we can encode as a plane partition of the desired type. Conversely, given such a plane partition, we can draw it as a stacking of cubes which in turn can be interpreted as a rhombic tiling.

An immediate corollary of the amazing correspondence between rhombic tilings and plane partitions is an explicit formula for the number $N(r,s,t)$ of rhombic tilings of $H(r,s,t)$. For this number is just the number of plane partitions with at most r rows, at most s columns, and with largest part at most t. If we set $x = 1$ in the left-hand side of MacMahon's formula (3.7) then it follows that we just get $N(r,s,t)$. If we set $x = 1$ in the right-hand side then we get the meaningless expression $0/0$. However, if we write

$$[i] = 1 - x^i = (1-x)(1 + x + \cdots + x^{i-1}),$$

then the factors of $1 - x$ cancel out from the numerator and denominator of the right-hand side of (3.7). Therefore substituting $x = 1$ is equivalent to replacing $[i]$ by the integer i, so we get the astonishing formula

$$N(r,s,t) = \frac{(1+t)(2+t)^2 \cdots (r+t)^r (r+1+t)^r \cdots (s+t)^r (s+1+t)^{r-1}}{1 \cdot 2^2 \cdot 3^3 \cdots r^r (r+1)^r \cdots s^r (s+1)^{r-1}} \cdot \frac{(s+2+t)^{r-2} \cdots (r+s-1+t)}{(s+2)^{r-2} \cdots (r+s-1)}.$$

10
The Möbius function and set partitions

This chapter introduces the Möbius function, a fundamental tool for enumerative combinatorics. It is very useful in situations demanding counting with "inclusion-exclusion," meaning that some objects get counted more than once and then the number of extra counts is subtracted to correct the error. There may possibly be several stages of corrections, and the pattern of adding and subtracting in such situations can be very complex. In some such cases, however, the Möbius function takes care of this bookkeeping for us, producing clean and computable answers.

Much can be said about the Möbius function, whose theory is very rich. Our aim here is however modest. We want to introduce this function and exemplify its use for an interesting and nontrivial enumeration problem, namely that of counting regions in space. This application is given in the next chapter.

The Möbius function

The *Möbius function* assigns a very significant integer to every finite *poset*. This word is an abbreviation for "partially ordered set." The precise definition of a poset is that it is a set P together with a relation $\leq$ satisfying the axioms:

- $x \leq x$ for all $x \in P$.

- If $x \leq y$ and $y \leq x$, then $x = y$.
- If $x \leq y$ and $y \leq z$, then $x \leq z$.

(Let us point out that it is possible for two elements x, y to satisfy neither $x \leq y$ nor $y \leq x$.) In the main examples that we will consider, the definition of $x \leq y$ has a simple, intuitive meaning such as "x is a subset of y" or "y is divisible by x." Thus it is not necessary to think in terms of the axiomatic definition above to understand our presentation.

For simplicity we assume that all posets considered have a bottom element less than all other elements and a top element greater than all other elements. Small posets can be represented as a diagram (graph) in which $x < y$ if there is a path from x to y moving up in the diagram. For instance, in Figure 18 we have $a < h$ since we can move up from a to h via d and g, for instance. Note that in this poset, a is the bottom element and h the top element.

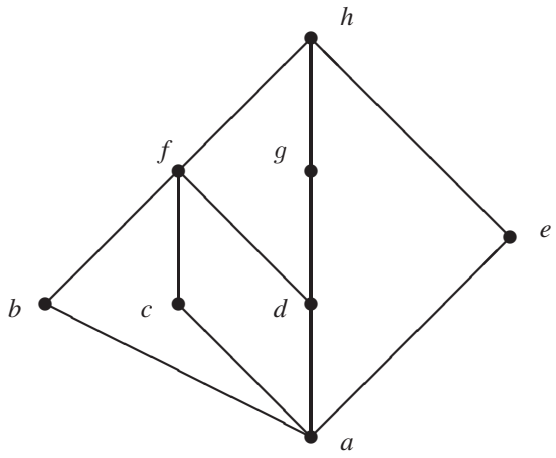


FIGURE 18
A small poset

The Möbius function $\mu(y)$ is recursively defined for any finite poset as follows: Put $\mu(a) = 1$ for the bottom element a of the poset, then require that

$$\mu(y) = -\sum_{x<y} \mu(x)$$

for all other elements y. This formula means that we are to define $\mu(y)$ so that when we sum $\mu(x)$ for all x less than or equal to y the resulting sum equals zero. This can clearly be done as long as one knows the values $\mu(x)$ for all elements x less than y.

To illustrate how this recursive definition works we compute the Möbius function of the poset in Figure 18, starting from the bottom. We get recursively:

$\mu(a) = 1$, by definition,
$\mu(b) = -\mu(a) = -1$,
$\mu(c) = -\mu(a) = -1$,
$\mu(d) = -\mu(a) = -1$,
$\mu(e) = -\mu(a) = -1$,
$\mu(f) = -\mu(a) - \mu(b) - \mu(c) - \mu(d) = -1 + 1 + 1 + 1 = 2$,
$\mu(g) = -\mu(a) - \mu(d) = -1 + 1 = 0$,
$\mu(h) = -\mu(a) - \mu(b) - \mu(c) - \mu(d) - \mu(e) - \mu(f) - \mu(g)$
$= -1 + 1 + 1 + 1 + 1 - 2 - 0 = 1$.

Figure 19 shows the same poset with computed Möbius function values.

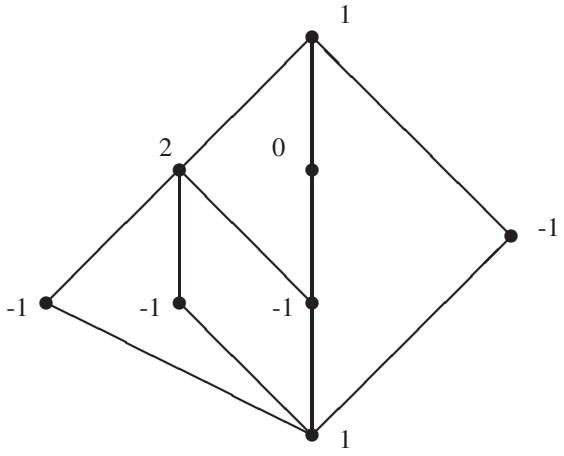


FIGURE 19
Values of the Möbius function

One important property of the Möbius function is that it can be used to "invert" summations over a partially ordered index set. Here is a statement

of the "Möbius inversion formula" in a special case. If a function $f : P \to \mathbf{Z}$ from a poset P to the integers is related to another function $g : P \to \mathbf{Z}$ by the partial summation formula

$$f(x) = \sum_{y \geq x} g(y),$$

then the value $g(a)$ at the bottom element a of P can be expressed in terms of f via the formula

$$g(a) = \sum_{y \in P} \mu(y) f(y).$$

The Möbius function has its origin in number theory, where it was introduced by August Ferdinand Möbius. (Möbius is best known to nonmathematicians for his eponymous connection with the "Möbius strip." The Möbius strip itself was well-known long before Möbius, but Möbius was one of the first persons to systematically investigate its mathematical properties.) The posets relevant to number theory are subsets of the positive integers ordered by divisibility. For these posets, the Möbius function $\mu(n)$ as defined above coincides with the classical Möbius function arising in number theory.

For instance, see the divisor diagram of the number 60 in Figure 20. The elements of this poset are the positive integers n that are divisors of 60, and we define $m \leq n$ in the poset if n is divisible by m. A calculation based on this diagram, analogous to the one we just carried out over Figure 18, shows that $\mu(60) = 0$. In the case of the classical Möbius function of number theory there is however a faster way to compute. Namely, for $n > 1$ one has that $\mu(n) = 0$ if the square of some prime number divides n, and that otherwise $\mu(n) = (-1)^k$ where k is the number of prime factors in n. Hence, for example: $\mu(60) = 0$ since $2^2 = 4$ divides 60; and $\mu(30) = -1$ since we have the prime factorization $30 = 2 \cdot 3 \cdot 5$ with an odd number of distinct prime factors.

The Möbius function is of great importance in number theory. Let it suffice to mention — for those who have the background to know what we are referring to — that both the Prime Number Theorem and the Riemann Hypothesis (considered by many to be the most important unsolved problem in all of mathematics) are equivalent to statements about the Möbius function. Namely, letting $M(n) = \sum_{k=1}^{n} \mu(k)$, it is known that

$$\text{Prime Number Theorem} \iff \lim_{n \to \infty} \frac{M(n)}{n} = 0,$$

$$\text{Riemann Hypothesis} \iff |M(n)| < n^{1/2+\epsilon}, \text{ for all } \epsilon > 0 \text{ and all sufficiently large } n.$$

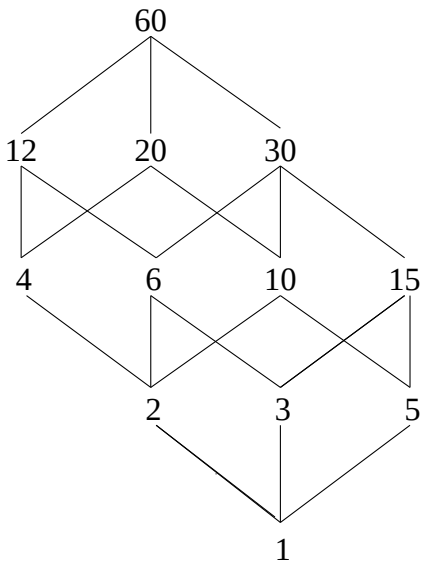


FIGURE 20
The divisors of 60

THE POSET OF SET PARTITIONS

Explicit expressions for the Möbius function have been determined for many posets of importance in combinatorics. In this section we present a basic example, namely a partial order on the partitions of a finite set.

A *partition of a set* A is a way of breaking it into smaller pieces, called *blocks*, namely a collection of pairwise disjoint nonempty subsets whose union is A. For instance, here are the 15 partitions of the set $\{1,2,3,4\}$:

1234, 12-34, 13-24, 14-23, 1-234, 2-134, 3-124, 4-123,
12-3-4, 13-2-4, 14-2-3, 23-1-4, 24-1-3, 34-1-2, 1-2-3-4

In the following we use $\{1,2,\ldots,n\}$ as the ground set and consider the collection of all partitions of this set. Denote this collection by Π_n.

There is a natural way to compare set partitions, saying that partition π is less than partition σ (written $\pi < \sigma$) if π is obtained from σ by further partitioning its blocks. This way we get a poset structure on the set Π_n, which

is illustrated in Figure 21 for Π_4. In accordance with the general definition of the diagram of a poset given earlier, this diagram is to be understood so that a partition π is less than a partition σ if and only if there is a downward path from σ to π in the order diagram, corresponding to further breaking up of σ's parts in order to obtain the partition π.

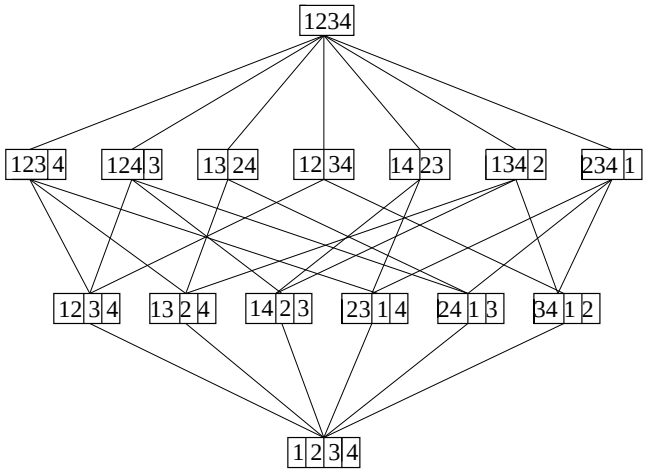


FIGURE 21
The poset Π_4 of set partitions.

Now we ask, what is the value μ_n of the Möbius function computed over the partition poset Π_n ? In other words, we want to determine $\mu_n = \mu_{\Pi_n}(\pi_{\max})$, where $\pi_{\max}$ denotes the top element of Π_n, the partition with only one block. For example, direct computation over the poset Π_4 in Figure 21 shows that $\mu_4 = -6$.

Here is the general answer:

$$\mu_n = (-1)^{n-1}(n-1)!. \tag{10.1}$$

To understand this formula it pays off to ask a more ambitious question. Let $b(\sigma)$ denote the number of blocks in a partition σ, and consider the polynomial

$$Q_n(x) \stackrel{\text{def}}{=} \sum_{\sigma \in \Pi_n} \mu(\sigma)\, x^{b(\sigma)} = \mu_n x + \text{higher-degree terms}, \tag{10.2}$$

which is a kind of generating function for the Möbius function of Π_n. What can be said about $Q_n(x)$?

Again, let us have a look at an example. Direct computation over the poset Π_4 in Figure 21 shows that

$$Q_4(x) = x^4 - 6x^3 + 11x^2 - 6x.$$

Observe that this polynomial $Q_4(x)$ is a product of some very simple polynomials of degree one:

$$Q_4(x) = x^4 - 6x^3 + 11x^2 - 6x = x(x-1)(x-2)(x-3).$$

This pretty factorization is not accidental. It is a special case of the general formula

$$Q_n(x) = x(x-1)(x-2)\cdots(x-(n-1)). \tag{10.3}$$

A proof for this will be outlined in the context of certain geometric arguments in the next chapter; see equations (11.4) and (11.5). In particular, as claimed by equation (10.1), we have that

$$\mu_n = (-1)(-2)\cdots(-(n-1)) = (-1)^{n-1}(n-1)!, \tag{10.4}$$

since $\mu_n x$ is the lowest-degree term of the polynomial $Q_n(x)$.

11
Arrangements and regions

Drawing a straight line in a plane splits the plane into two parts. Drawing several lines cuts it into several parts. How many? This kind of geometric partitioning problem is discussed here. A moments thought makes it clear that the answer depends not only on the number of lines, but also on how they intersect. How, if at all, can the dependence be described?

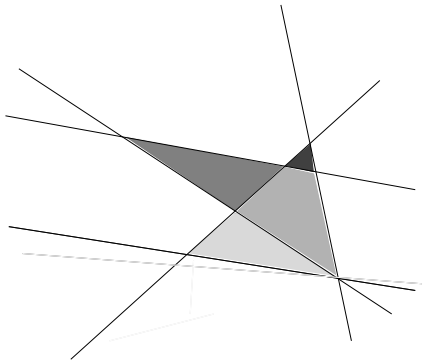


FIGURE 22
An arrangement of 5 lines.

The analogous question can be asked also for higher dimensions, and it turns out that there is an elegant general answer involving the Möbius function.

ARRANGEMENTS OF LINES

A collection $\mathcal{A}$ of straight lines in the plane is called a *line arrangement*, and the connected pieces into which the lines partition the plane are called *regions*. There are two kinds of regions, the *bounded* ones (which are contained inside some circle) and the *unbounded* ones (running off to infinity). How many regions are there? Of these, how many are bounded? For instance, the arrangement in Figure 22 has 14 regions, of which 4 are bounded (the shaded ones).

Suppose that the arrangement $\mathcal{A}$ consists of t lines, and let $r(\mathcal{A})$ and $b(\mathcal{A})$ denote the number of regions and the number of bounded regions, respectively, into which the plane is divided by $\mathcal{A}$.

If all the lines are parallel we see immediately that $r(\mathcal{A}) = t + 1$ and $b(\mathcal{A}) = 0$. So, let us assume that there is at least one intersection point where two lines meet. Let C denote the set of intersection points, and for each point p in C let m_p denote the number of lines passing through p.

Theorem. *We have that*

(i)
$$r(\mathcal{A}) = 1 + t + \sum_{p \in C} (m_p - 1),$$

(ii)
$$b(\mathcal{A}) = 1 - t + \sum_{p \in C} (m_p - 1).$$

For instance, the arrangement in Figure 22 has $t = 5$ lines, six intersection points with $m_P = 2$ and one intersection point with $m_P = 3$; see Figure 23. (Remark: Disregard for now the labels on points and lines in Figure 23 — they will be of use later on.) Hence the equations in the theorem produce the correct answers $r(\mathcal{A}) = 1 + 5 + 6 \cdot 1 + 1 \cdot 2 = 14$, and $b(\mathcal{A}) = 1 - 5 + 6 \cdot 1 + 1 \cdot 2 = 4$.

Two more examples of line arrangements are given in Figures 28 and 29. The reader is invited to compute their numbers of regions using equations (i) and (ii).

We sketch a proof of the theorem by so-called "mathematical induction." A proof by induction works like this: one verifies the statement for some particular value, in our case $t = 2$, and then one proves that the truth of the statement for some value $t - 1$ implies its truth for the value t. Then the

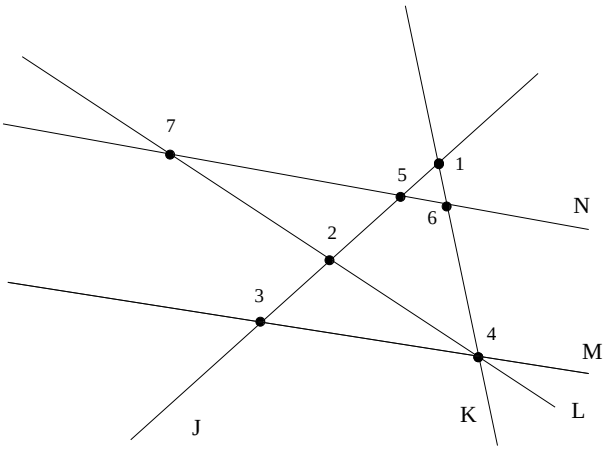


FIGURE 23
The seven intersection points.

truth for $t = 2$ implies the truth for $t = 3$, which in turn implies the truth for $t = 4$, and so on, so that the truth for all $t \geq 2$ will follow.

Plugging $t = 2$ into equations (i) and (ii) we get $r(\mathcal{A}) = 4$ and $b(\mathcal{A}) = 0$, which is certainly correct for any arrangement consisting of two crossing lines.

Now, assume that the theorem is correct for all arrangements having $t-1$ lines. We want to prove it for an arrangement $\mathcal{A}$ having t lines. Then argue as follows. Remove one of the lines, call it L, from $\mathcal{A}$. This leaves a smaller arrangement $\mathcal{A}'$ with only $t-1$ lines. How many new regions are created when we reintroduce the line L ?

This can be seen in the following way. Imagine that we travel along the line L, coming in from infinity and after crossing some of the lines of $\mathcal{A}'$ heading out to infinity again in the opposite direction; see Figure 24 where L is dashed.

Traversing L we encounter some points where L intersects other lines. Let C_L denote the set of such intersection points and suppose that there are k of them. Then each time we reach a crossing point we complete the splitting of the region of $\mathcal{A}'$ that we have just traversed. Finally, after the last crossing point is reached we end by splitting the one remaining region. Thus, we have created $k+1$ new regions in all.

Since the arrangement $\mathcal{A}'$ has only $t-1$ lines, we get by the induction assumption that

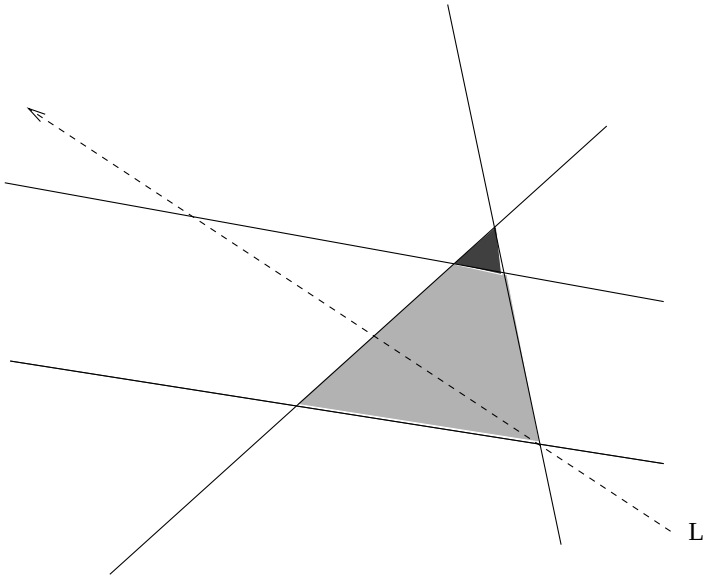


FIGURE 24
Traversing the line L through the subarrangement $\mathcal{A}'$.

$$r(\mathcal{A}') \quad = \quad 1 + (t-1) + \sum_{p \in C \setminus C_L} (m_p - 1) + \sum_{p \in C_L} (m_p - 2),$$

where $C \setminus C_L = \{p \in C : p \notin C_L\}$. Hence, since there are $k+1$ new regions and $C_L \subseteq C$, we obtain

$$\begin{aligned} r(\mathcal{A}) &= r(\mathcal{A}') + k + 1 \\ &= 1 + (t-1) + \sum_{p \in C \setminus C_L} (m_p - 1) + \sum_{p \in C_L} (m_p - 2) + k + 1 \\ &= 1 + (t-1) + \sum_{p \in C \setminus C_L} (m_p - 1) + \sum_{p \in C_L} (m_p - 1) - k + k + 1 \\ &= 1 + t + \sum_{p \in C} (m_p - 1). \end{aligned}$$

Thus we have verified equation (i). Equation (ii) can be proved by similar reasoning, observing that exactly $k-1$ of the newly created regions are bounded (all but the first and the last along L). Another way to reason is that equation (ii) follows from equation (i) since there must be exactly $2t$ unbounded regions. To see this, imagine a circle drawn in the plane and large

enough that all intersection points of $\mathcal{A}$ are in its interior. Then the unbounded regions are precisely the regions that reach outside the circle, and since the lines don't intersect outside the circle it is easy to see by going once around the circle that they cut out precisely $2t$ regions there.

Arrangements of hyperplanes

Can this kind of reasoning be generalized from 2-dimensional space (the plane) to 3-dimensional space, or even to higher n-dimensional space? Let us consider for a moment the case of 3 dimensions. The analogue of a line is here a plane. Mathematically planes can be described by equations of the form $ax+by+cz=d$ in three space coordinates x, y and z. Familiarity with linear algebra is of course helpful, but it should suffice for the following to have an idea of the intuitive notion of a plane in ordinary 3-space.

A plane cuts space into two regions, several planes cut space into several regions. As before we ask: how many? And, how many of these regions are bounded? The reader is invited to try to carry out an argument similar to the one we have given for line arrangements. However, this kind of barehanded approach is tricky in the 3-dimensional case. Fortunately, to our rescue comes the Möbius function.

Let us consider a finite collection $\mathcal{A} = \{H_1, \ldots, H_t\}$ of hyperplanes H_i in n-dimensional space $\mathbf{R}^n$. By "hyperplane" is here meant the solution set to a linear equation $a_1x_1+\cdots+a_nx_n = b$. A reader who is not comfortable with the notion of higher-dimensional space can think about our usual 3-dimensional space, where "hyperplanes" are just ordinary planes.

The arrangement $\mathcal{A}$ cuts space into a number of connected components, called *regions*, some of which are bounded. Let $r(\mathcal{A})$ and $b(\mathcal{A})$ denote the number of regions and the number of bounded regions, respectively.

The theorem on page 92 shows that for lines in the plane the numbers $r(\mathcal{A})$ and $b(\mathcal{A})$ can be computed from information about how the lines intersect. It turns out that something similar is true in higher dimensions To describe the intersection pattern we introduce the *intersection poset* $L_{\mathcal{A}}$. This is a poset consisting of all nonempty intersections $\cap_{H\in\mathcal{B}}H$ of subfamilies $\mathcal{B}\subseteq\mathcal{A}$, and also the whole space $\mathbf{R}^n$ itself. The partial order is the opposite of set inclusion: $\sigma \leq \pi$ in $L_{\mathcal{A}}$ if and only if $\pi \subseteq \sigma$. In particular, $\mathbf{R}^n$ is the bottom element. Figure 25 shows the intersection poset of the line arrangement of Figure 23.

It is a remarkable fact that the numbers of regions of $\mathcal{A}$ can be computed from the values of the Möbius function of $L_{\mathcal{A}}$, as shown by the following

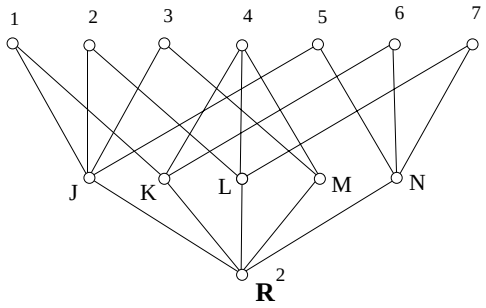


FIGURE 25
Intersection poset.

theorem proved by Thomas Zaslavsky in 1975.

Zaslavsky's Theorem. *We have that*

$$r(\mathcal{A}) = \sum_{\sigma \in L_{\mathcal{A}}} \mid \mu(\sigma) \mid \quad \textit{and} \quad b(\mathcal{A}) = \mid \sum_{\sigma \in L_{\mathcal{A}}} \mu(\sigma) \mid$$

(*Remark:* The formula for $b(\mathcal{A})$ requires that the arrangement $\mathcal{A}$ is *essential*, meaning that the normals of the hyperplanes span $\mathbf{R}^n$.)

We have already proved Zaslavsky's theorem for dimension two, because for line arrangements in the plane the theorem specializes to the theorem on page 92. For an example, take the line arrangement in Figure 22. Its intersection poset is shown in Figure 25 and the Möbius function values are shown in Figure 26.

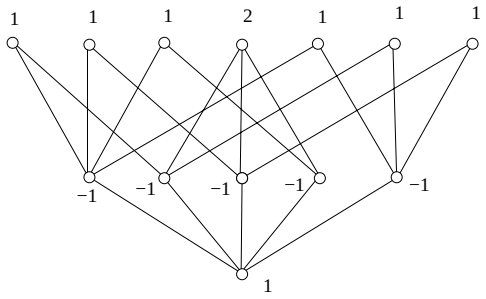


FIGURE 26
Möbius function values of intersection poset.

A convenient way of collecting the relevant information about the Möbius function of the intersection poset of an arrangement $\mathcal{A}$ is via its *characteristic polynomial*, defined as follows:

$$P_{\mathcal{A}}(x) \stackrel{\text{def}}{=} \sum_{\sigma \in L_{\mathcal{A}}} \mu(\sigma)\, x^{\dim(\sigma)} = x^n + \text{lower-degree terms.} \tag{11.1}$$

For instance, for the arrangement $\mathcal{A}$ in Figures 22 and 26 we have that

$$P_{\mathcal{A}}(x) = x^2 - 5x + 8.$$

It can be shown that $(-1)^{n-\dim(\sigma)}\mu(\sigma) > 0$ for all $\sigma \in L_{\mathcal{A}}$. Zaslavsky's theorem can therefore be reformulated in the following compact way:

$$r(\mathcal{A}) = |\, P_{\mathcal{A}}(-1)\,| \tag{11.2}$$

$$b(\mathcal{A}) = |\, P_{\mathcal{A}}(1)\,| \tag{11.3}$$

Thus the characteristic polynomial carries the enumerative information about regions, bounded and general, in which we are interested. But, how can one compute this polynomial? Doesn't its determination require that we somehow compute all Möbius function values, in which case no simplification is achieved, only a reformulation. Or are there shortcuts to computing characteristic polynomials?

Yes, there are, but only for *some* arrangements, namely for those whose hyperplanes are given by linear equations $a_1x_1 + \cdots + a_nx_n = b$ with *integer* coefficients a_i and b. Then an interesting method is available that in some cases allows elegant and easy computations.

Theorem. *Let $\mathcal{A}$ be a hyperplane arrangement in $\mathbf{R}^n$, determined by equations $a_1x_1 + \cdots + a_nx_n = b$ with integer coefficients a_i and b. Then, for all sufficiently large*[1]*) prime numbers p, the value $P_{\mathcal{A}}(p)$ of the characteristic polynomial (i.e., $P_{\mathcal{A}}(x)$ evaluated at $x = p$) equals the number of n-tuples of integers $(q_1, q_2, \ldots, q_n)$ such that*

(1) $0 \le q_i \le p-1$, for all i

(2) the integer $a_1q_1 + \cdots + a_nq_n - b$ is not divisible by p, for all the equations $a_1x_1 + \cdots + a_nx_n = b$ that determine $\mathcal{A}$.

It is a consequence of the theorem that in order to count regions of such an arrangement in real n-dimensional space we can instead count certain n-tuples of integers relative to large prime numbers p. This is so because of

[1]) The meaning of saying that p has to be "sufficiently large" can be made precise. In particular, there are finitely many "bad" primes that must be avoided.

the fact that a polynomial is determined by its evaluation at sufficiently many points, in this case at the infinitely many (sufficiently large) prime numbers p.

The theorem gives rise to a method for enumerating regions of an arrangement $\mathcal{A}$ via its characteristic poynomial. The first step of the method is to express $P_{\mathcal{A}}(p)$ as a function of p by counting n-tuples of numbers as described in the theorem. Then we forget that p is a prime number and think of p as a variable, and finally we make the substitutions $p = -1$ and $p = 1$. Then by formulas (11.2) and (11.3) we will have computed the number of regions and the number of bounded regions.

This method will be illustrated in the next three sections.

THE BRAID ARRANGEMENT AND SET PARTITIONS

Consider the hyperplane arrangement $\mathcal{B}_n$ in $\mathbf{R}^n$, called the *braid arrangement*, given by the equations

$$x_i - x_j = 0, \text{ for } 1 \leq i < j \leq n.$$

In spite of its simple appearance, this arrangement plays an important role at the intersection of several areas of mathematics: algebra, topology, combinatorics, and more. How many regions are there in its complement? This is easy to answer. The following direct argument shows that there are $n!$ regions.

In order to specify which region R a point $a = (a_1, \ldots, a_n)$ in $\mathbf{R}^n$ belongs to, we must describe on which side of each hyperplane $x_i - x_j = 0$ does a lie. In other words, for all $1 \leq i < j \leq n$, we must specify whether $a_i < a_j$ or $a_i > a_j$. This is the same as specifying a *linear ordering* of $a_1, \ldots, a_n$, i.e., a permutation π of $1, 2, \ldots, n$ such that $a_{\pi(1)} < a_{\pi(2)} < \cdots < a_{\pi(n)}$. Since there are $n!$ permutations π, it follows that there are $n!$ regions R_π, defined by

$$R_\pi = \{(a_1, a_2, \ldots, a_n) \in \mathbf{R}^n \,:\, a_{\pi(1)} < a_{\pi(2)} < \cdots < a_{\pi(n)}\}.$$

Let us test the machinery presented in this chapter by computing the number of regions of $\mathcal{B}_n$ via its characteristic polynomial. The question then is: How many n-tuples $(q_1, \ldots, q_n)$ of numbers q_i are there such that $0 \leq q_i \leq p-1$ for all i and $q_i \neq q_j$ for all $1 \leq i < j \leq n$? This counting problem is easy. There are p ways to choose q_1. After that there are $p-1$ ways to choose q_2 (all choices are valid except the value already chosen for q_1). Then there are $p-2$ ways to choose q_3, and so on. We find that there are

$$p(p-1)(p-2)\cdots(p-n+1)$$

choices in all. (Remember that we need only consider large prime numbers p, in particular we may assume that $p > n+1$.) By the theorem on page 97 we then obtain the characteristic polynomial

$$P_{\mathcal{B}_n}(x) = x(x-1)(x-2)\cdots(x-n+1). \tag{11.4}$$

Plugging in $x = -1$ we get the correct number of regions

$$r(\mathcal{B}_n) = |\,(-1)(-2)(-3)\cdots(-n)\,| = n!$$

The intersection poset of the braid arrangement has a very concrete combinatorial description. Consider, for example, what it means for a point $(a_1, \ldots, a_7) \in \mathbf{R}^7$ to lie in the intersection of the three hyperplanes $x_2 - x_4 = 0$, $x_3 - x_5 = 0$ and $x_4 - x_7 = 0$ This happens if $a_2 = a_4$, $a_3 = a_5$ and $a_4 = a_7$. In other words, we require precisely that $a_2 = a_4 = a_7$ and $a_3 = a_5$, and otherwise the real numbers a_i can be arbitrary. All the information is therefore encoded in the family of sets $\{1\}$, $\{2,4,7\}$, $\{3,5\}$, $\{6\}$, indicating which groups of variables should be set equal. This family of sets is, of course, nothing other than a partition of the set $\{1, 2, \ldots, 7\}$.

The general description can be gleaned from this special case. The elements of Π_n, i.e. set partitions, encode the intersections of subfamililes of hyperplanes from $\mathcal{B}_n$. Furthermore, such an intersection X_π properly contains another intersection X_σ if and only if the corresponding partition π is less than σ in the poset Π_n. In other words,

Π_n is the intersection poset of the braid arrangement $\mathcal{B}_n$.

Recall the notation $b(\sigma)$ for the number of blocks in a partition σ. The dimension of the subspace X_π equals the number of blocks in the partition π : $\dim(X_\pi) = b(\pi)$. Hence, recalling the definitions (10.2) and (11.1), we conclude that the characteristic polynomial of $\mathcal{B}_n$ equals the polynomial $Q_n(x)$:

$$P_{\mathcal{B}_n}(x) = Q_n(x). \tag{11.5}$$

Equation (10.3) follows from this and equation (11.4).

Fattened braid arrangements

Here we consider what happens to the number of regions if the braid arrangement is "fattened up" by adding parallel hyperplanes. Consider the arrangement $\mathcal{B}_n^k$ in $\mathbf{R}^n$ defined by the equations

$$x_i - x_j \in \{0, \pm 1, \pm 2, \ldots, \pm k\}, \text{ for all } 1 \le i < j \le n.$$

Thus, there are $(2k+1)\binom{n}{2}$ defining equations. For $k=0$ we have the braid arrangement, and in general $\mathcal{B}_n^k$ is obtained from $\mathcal{B}_n^0 = \mathcal{B}_n$ by adding k parallel hyperplanes on each side of the $\binom{n}{2}$ original hyperplanes. Figures 27 and 28 show the arrangements $\mathcal{B}_3^0$ and $\mathcal{B}_3^2$. The figures actually depict the intersection of these arranagements with the plane $x_1+x_2+x_3=0$, and hence show 2-dimensional line arrangements — the motivation for this is discussed at the end of this section. The actual arrangements $\mathcal{B}_3^0$ and $\mathcal{B}_3^2$ are obtained by replacing each line L with a plane through L perpendicular to the diagram. This procedure does not affect the intersection poset or the number of regions, as explained in more detail in the remark at the end of this chapter.

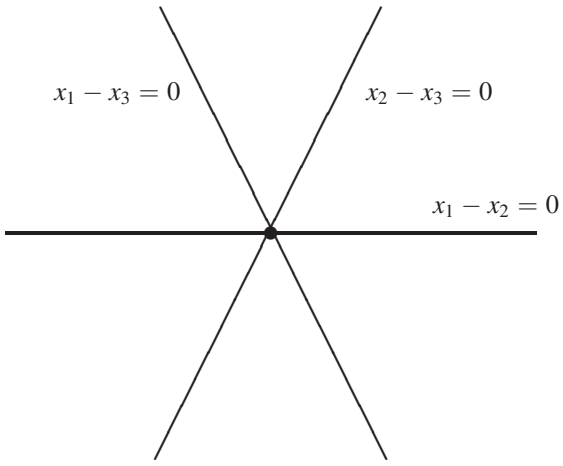


FIGURE 27
The braid arrangement $\mathcal{B}_3 = \mathcal{B}_3^0$ intersected with $x_1 + x_2 + x_3 = 0$.

How many regions does $\mathcal{B}_n^k$ have? We count n-tuples $(q_1, q_2, \ldots, q_n)$ with $0 \le q_i \le p-1$ such that

$$q_i - q_j \notin \{0, \pm 1, \ldots, \pm k\}, \text{ for all } 1 \le i < j \le n.$$

In other words, we want to count n-tuples $(q_1, q_2, \ldots, q_n)$ such that

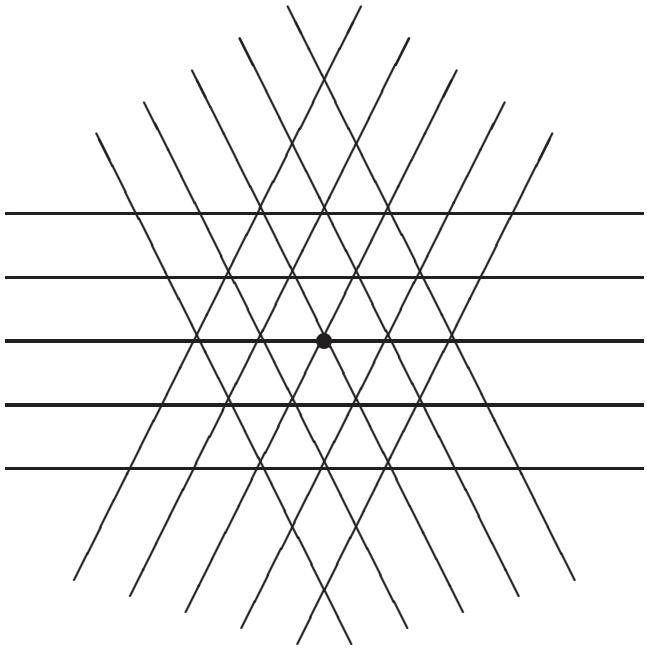


FIGURE 28
The arrangement $\mathcal{B}_3^2$ intersected with $x_1 + x_2 + x_3 = 0$

$$q_i \in \{0, 1, 2, \ldots, p-1\}, \text{ for all } i = 1, 2, \ldots, n, \tag{11.6}$$

$$\text{the } q_i \text{ are distinct,} \tag{11.7}$$

$$q_i - q_j \neq s, \text{ for all } 1 \leq i < j \leq n \text{ and all } s \in \{\pm 1, \ldots, \pm k\}. \tag{11.8}$$

This counting problem has the following reformulation.

Say that we have a round table surrounded by p chairs numbered consecutively $0, 1, 2, \ldots, p-1$. Suppose that a group of n people (with $n < p$) is to be seated around the table in these chairs. We can call these people $X_1, X_2, \ldots X_n$.

Let us say that $q_i = r$ means that person X_i sits in chair number r. Then conditions (11.6) and (11.7) mean that they all sit in separate chairs. Condition (11.8) has the following meaning. Since $q_i \neq q_j + s$ for $s = 1, 2, \ldots, k$, X_i is not allowed to sit in any of the k chairs immediately to the right of X_j, and since $q_i \neq q_j - s$ for $s = 1, 2, \ldots, k$, X_i is not allowed to sit in any of the k chairs immediately to the left of X_j. In other words, there must be k empty chairs between any two people sitting at the table.

Thus, the enumerative problem we want to solve is the following:

In how many ways can n people be seated in p chairs around a table so that there are at least k empty chairs between any two of them?

A small argument, given in the appendix to this chapter, gives the following.

Answer: in $p(p-kn-1)(p-kn-2)\cdots(p-kn-n+1)$ ways.

From this we deduce the characteristic polynomial

$$P_{\mathcal{B}_n^k}(x) = x(x-kn-1)(x-kn-2)\cdots(x-kn-n+1),$$

and by setting $x=-1$ we obtain the number of regions.

$$r(\mathcal{B}_n^k) = (nk+2)(nk+3)\cdots(nk+n) = \frac{(nk+n)!}{(nk+1)!}$$

For instance, the arrangement $\mathcal{B}_3^3$ of 21 planes in $\mathbf{R}^3$ has 132 regions.

The following remark is intended for readers with a sufficient knowledge of geometry. All hyperplanes of the arrangement $\mathcal{B}_n^k$ are orthogonal to the hyperplane H_0 given by the equation $x_1+x_2+\cdots+x_n=0$. We obtain another arrangement, call it $\widehat{\mathcal{B}}_n^k$, by intersecting the hyperplanes of $\mathcal{B}_n^k$ with H_0. This gives an essentially equivalent arrangement in the $(n-1)$-dimensional space H_0, in which all dimensions are reduced by one.

One can easily prove that the intersection poset of $\widehat{\mathcal{B}}_n^k$ is isomorphic to that of $\mathcal{B}_n^k$. The dimension of an intersection is, however, one less in $\widehat{\mathcal{B}}_n^k$. Therefore the characteristic polynomial of $\widehat{\mathcal{B}}_n^k$ can be obtained simply by dividing that of $\mathcal{B}_n^k$ by x. Thus,

$$P_{\widehat{\mathcal{B}}_n^k}(x) = (x-kn-1)(x-kn-2)\cdots(x-kn-n+1).$$

From this we obtain the number of regions and the number of bounded regions by substituting $-x=1$ and $x=1$, respectively:

$$r(\widehat{\mathcal{B}}_n^k) = (nk+2)(nk+3)\cdots(nk+n) = \frac{(nk+n)!}{(nk+1)!}$$

and

$$b(\widehat{\mathcal{B}}_n^k) = (nk)(nk+1)\cdots(nk+n-2) = \frac{(nk+n-2)!}{(nk-1)!}.$$

For instance, the line arrangement $\widehat{\mathcal{B}}_3^2$ of 15 lines in $\mathbf{R}^2$ has 72 regions, of which 42 are bounded. This can also be seen from Figure 28.

SHI ARRANGEMENTS AND LABELED TREES

Define the arrangement $\mathcal{C}_n$ in $\mathbf{R}^n$ by the equations

$$x_i = 0 \text{ and } x_i = 1, \text{ for all } i = 1, 2, \ldots, n,$$

and

$$x_i - x_j = 0 \text{ and } x_i - x_j = 1, \text{ for all } 1 \leq i < j \leq n.$$

Figure 29 shows the arrangement $\mathcal{C}_2$. These arrangements are equivalent to the so called *Shi arrangements* arising in the study of affine Weyl groups. Notice that $\mathcal{C}_n$ contains the braid arrangement $\mathcal{B}_n$ as a subarrangement.

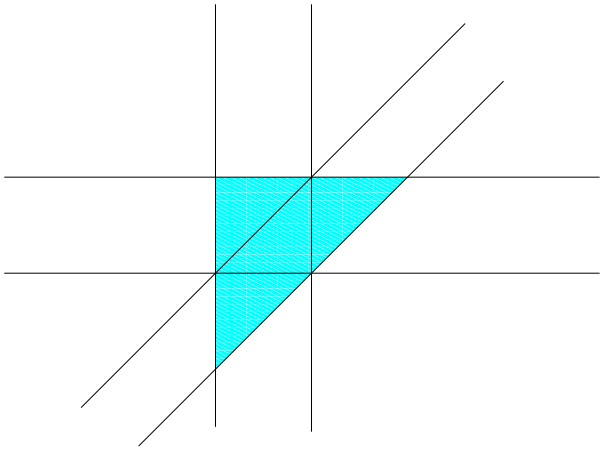


FIGURE 29
The arrangement $\mathcal{C}_2$.

How many regions does $\mathcal{C}_n$ have? According to the theorem on page 97 we should count n-tuples $(q_1, q_2, \ldots, q_n)$ with $0 \leq q_i \leq p-1$ such that

$$q_i \neq 0, 1, \text{ for all } i = 1, 2, \ldots, n,$$

and

$$q_i - q_j \neq 0, 1, \text{ for all } 1 \leq i < j \leq n.$$

In other words, we want to count n-tuples $(q_1, q_2, \ldots, q_n)$ such that

$$q_i \in \{2, 3, \ldots, p-1\}, \text{ for all } i = 1, 2, \ldots, n, \tag{11.9}$$

$$\text{the } q_i \text{ are distinct,} \tag{11.10}$$

$$q_i - q_j \neq 1, \text{ for all } 1 \leq i < j \leq n. \tag{11.11}$$

This counting problem can be dealt with using the following model. Say that we have a long table with chairs along one side numbered consecutively $2, 3, \ldots, p-1$. A group of n people of different ages is to be seated in these chairs. We call these people $X_1, X_2, \ldots X_n$, and assume that they are listed in order of increasing age: X_1 is the youngest, then comes X_2, etc, so that X_n is the oldest.

Let us say that $q_i = r$ means that person X_i sits in chair number r. Then conditions (11.9) and (11.10) mean simply that there is at most one person in each of the $p-2$ chairs (no one is sitting in someone else's lap). Condition (11.11), $q_i \neq q_j + 1$, is more interesting. It means that if X_i is younger than X_j, then X_i is not allowed to sit in the chair immediately to the right of X_j.

Thus, the enumerative problem we want to solve is the following.

In how many ways can n people of different ages be seated in $p-2$ linearly arranged chairs so that no one has a younger person immediately to his/her right?

A small argument, given in the appendix to this chapter, gives the following.

Answer: In $(p-n-1)^n$ ways.

We have this way, via the theorem on page 97, computed the characteristic polynomial $P_{\mathcal{C}_n}(x) = (x-n-1)^n$, and by setting $x = -1$ and $x = 1$ we can deduce the number of regions and the number of bounded regions, respectively:

$$r(\mathcal{C}_n) = (n+2)^n \quad \text{and} \quad b(\mathcal{C}_n) = n^n. \tag{11.12}$$

For instance, the line arrangement $\mathcal{C}_2$ has sixteen regions, four of which are bounded; see Figure 29.

Does the number $(n+2)^n$ ring a bell? In Chapter 1 we enumerated the labeled trees on $n+2$ vertices and came up with exactly this number (see Cayley's theorem, p. 13). Is there any connection?

It would take us too far afield to continue this discussion here, but the answer is yes: explicit and natural bijections between the set of regions of $\mathcal{C}_n$ and the set of labeled trees on $n+2$ vertices have been described. Via such bijections one can obtain an alternative, bijective proof of the formula (11.12) for $r(\mathcal{C}_n)$.

APPENDIX: COUNTING TABLE SEATING PLACEMENTS

1. Let $U(n,p,k)$ denote the number of ways in which n people can be seated in p chairs around a table so that there are at least k empty chairs between any two of them. To determine $U(n,p,k)$ we need the following.

Observation. *There are*

$$\binom{p-k(m-1)}{m}$$

ways to choose an m-set $a_1 < a_2 < \cdots < a_m$ from $\{1,2,\ldots,p\}$ so that $a_{i+1} - a_i > k$ for all i.

Proof. For any such sequence $a_1 < a_2 < \cdots < a_m$, let $b_i = a_i - k(i-1)$. Then

$$b_{i+1} - b_i = a_{i+1} - a_i - k.$$

Hence, $a_{i+1} - a_i > k$ if and only if $b_{i+1} - b_i > 0$. We may therefore instead count sequences $b_1 < b_2 < \cdots < b_m$ from $\{1,2,\ldots,p-k(m-1)\}$ such that $b_{i+1} - b_i > 0$ for all i. This is the same as counting m-element subsets of $\{1,2,\ldots,p-k(m-1)\}$, a number given by the binomial coefficient, completing the proof of the observation.

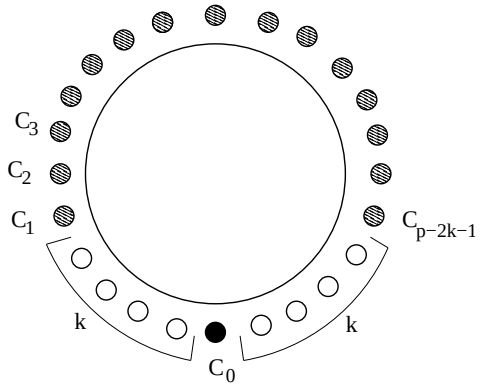


FIGURE 30
The seating arrangement: choosing among the grey chairs

Now we return to the determination of $U(n,p,k)$. The first person chooses a chair C_0 and sits down. This can be done in p ways, since there are p

chairs around the table. Now the chair C_0 and the k chairs to its immediate right as well as the k chairs to its immediate left are "forbidden," so there are $p-(1+2k)$ chairs left for the others to choose. Call these chairs $C_1,\ldots,C_{p-2k-1}$, listed in their clockwise order. They are shaded grey in Figure 30.

A moment's thought shows that a sequence of chairs C_{a_i} is allowed to be chosen if and only if the sequence of indices $a_1<a_2<\cdots<a_{n-1}$ satisfies the requirement that $a_{i+1}-a_i>k$ for all i. Therefore, the observation above gives the number of ways that the $n-1$ remaining chairs can be chosen. Finally, there are $(n-1)!$ ways that the $n-1$ people can be distributed to the chosen chairs, so in all we get

$$\begin{aligned} U(n,p,k) &= p\,(n-1)!\binom{(p-2k-1)-k\,(n-2)}{n-1} \\ &= p\,(n-1)!\,\frac{(p-kn-1)!}{(n-1)!\,(p-kn-n)!} \\ &= p(p-kn-1)(p-kn-2)\cdots(p-kn-n+1). \end{aligned}$$

2. Let $W(n,p)$ denote the number of ways in which n people of different ages can be seated in $p-2$ chairs, linearly arranged along one side of a table, so that no one has a younger person immediately to his/her right.

Suppose we have such a seating arrangement. Then n of the $p-2$ chairs are taken, so $p-n-2$ empty ones remain. The empty chairs partition the set N of people into $p-n-1$ groups, namely the group N_1 of people sitting to the left of the first empty chair, the group N_2 sitting between the first and second empty chairs, and so on until the last group N_{p-n-1} sitting to the right of the last empty chair. Notice that some of these groups may be empty, which happens when there are two empty chairs next to each other.

The sequence $N_1,N_2,\ldots,N_{p-n-1}$ of groups, is a *weak ordered partition* of the set N of n people, by which we mean an ordered sequence of pairwise disjoint subsets N_i (empty set allowed, hence the term "weak") whose union is N.

The crucial observation is that the seating arrangement is completely determined by the weak ordered partition, since each group N_i of people must sit in order of increasing age and the positions of the empty chairs are determined. Hence, we want to know: how many weak ordered partitions of N with $p-n-1$ blocks are there?

Think of it this way. To create a weak ordered partition with $p-n-1$ blocks, each of the n people must be placed in one of $p-n-1$ groups, with no further restrictions. For each person there are $p-n-1$ choices, each choice independent of the other ones. Hence, for the total number of choices

$$W(n,p) = (p-n-1)^n$$

12
Face numbers of polytopes

Among the many results of Euler that have initiated fruitful lines of development in combinatorics, the one that is perhaps most widely known is his famous counting formula for boundaries of 3-dimensional polytopes from 1752. Here we trace parts of a story that spans a couple of centuries and involves several branches of mathematics.

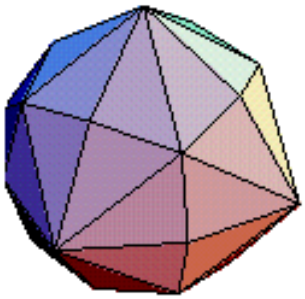


FIGURE 31
A 3-dimensional convex polytope.

EULER'S FORMULA

A 3-*polytope* P (or 3-dimensional convex polytope, to be more precise) is a bounded region of space obtained as the intersection of finitely many halfspaces (and not contained in any plane). It is precisely the kind of geometric object that can arise as a "bounded region" of an arrangement of planes in 3-dimensional space, in the sense explained in Chapter 11. For the layman it can be described as the kind of solid body you can create from a block of cheese with a finite number of plane cuts with a knife.

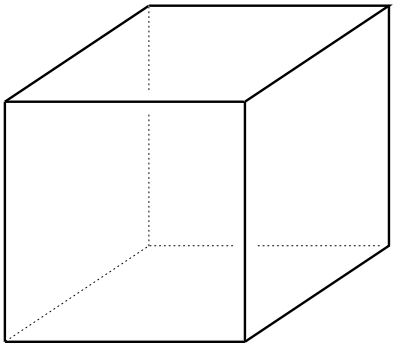


FIGURE 32
The cube.

For instance, take the ordinary cube shown in Figure 32 — it can be cut out with six plane cuts. The cube is one of the five *Platonic solids: tetrahedron, cube, octahedron, dodecahedron and icosahedron*, known and revered by the Greek mathematicians in antiquity. A more general example of a 3-polytope is shown in Figure 31.

A polytope that is dear to all combinatorialists is the "permutohedron," shown in Figure 33. Its 24 corners correspond to the $24 = 4{\cdot}3{\cdot}2{\cdot}1$ permutations of the set $\{1, 2, 3, 4\}$. The precise rule for constructing the permutohedron and for labelling its vertices with permutations is best explained in 4-dimensional space and will be left aside. Note that the pairs of permutations that correspond to edges of the permutohedron are precisely pairs that differ by a switch of two adjacent entries, such as 2143 — 2134 or 3124 — 3214. Thus, edge-paths on the boundary of the permutohedron are precisely paths consisting of such "adjacent transpositions," giving geometric content to the topic of reduced decompositions that was discussed in Chapter 7.

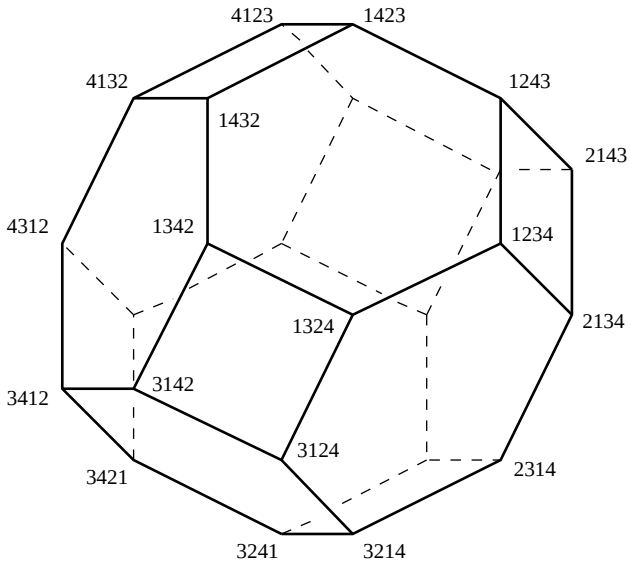


FIGURE 33
The permutohedron.

The boundary of a 3-polytope is made up of pieces of dimension 0, 1 and 2 called its *faces*. These are the possible areas of contact if the polytope is made to touch a plane surface, such as the top of a table. The 0-faces are the corners, or vertices, of the polytope. The 1-faces are the edges, and the 2-faces are the flat surfaces, such as the six squares bounding the cube. The permutohedron has fourteen 2-faces, six of which are 4-sided and eight are 6-sided.

Euler's formula has to do with counting the number of faces of dimensions 0, 1 and 2. Namely, let f_i be the number of i-dimensional faces.

Euler's Formula. *For any* 3*-polytope:*

$$f_0 - f_1 + f_2 = 2.$$

Let us verify this relation for the cube and the permutohedron:

	f_0	f_1	f_2	$f_0 - f_1 + f_2$
Cube	8	12	6	$8 - 12 + 6 = 2$
Permutohedron	24	36	14	$24 - 36 + 14 = 2$

From a modern mathematical point of view there is no difficulty in defining higher-dimensional polytopes. Thus, a *d-polytope* is a full-dimensional bounded intersection of closed halfspaces in $\mathbf{R}^d$. Such higher-dimensional polytopes have taken on great practical significance in the last fifty years due to their importance for linear programming. The term "linear programming" refers to techniques for optimizing a linear function subject to a collection of linear constraints. The linear constraints cut out a feasible region of space, which is a d-polytope (possibly unbounded in this case). The combinatorial study of the structure of polytopes has interacted very fruitfully with this applied area.

It can be shown that the same definition of the *faces* of a polytope works also in higher dimensions (namely, "the possible areas of contact if the polytope is made to touch a $(d-1)$-dimensional hyperplane in $\mathbf{R}^d$"), and that there are only finitely many faces of each dimension $0, 1, \ldots, d-1$. Thus we may define the number f_i of i-dimensional faces for $i = 0, 1, \ldots, d-1$. These numbers for a given polytope P are collected into a string

$$f(P) = (f_0, f_1, \ldots, f_{d-1}),$$

called the *f-vector* of P. For instance, we have seen that $f(\text{cube}) = (8, 12, 6)$ and $f(\text{permutohedron}) = (24, 36, 14)$.

Is there an Euler formula for f-vectors in higher dimensions? This question was asked early on, and by the mid-1800's some mathematicians had discovered the following beautiful fact.

Generalized Euler Formula. *For any d-polytope:*

$$f_0 - f_1 + f_2 - \cdots + (-1)^{d-1} f_{d-1} = 1 + (-1)^{d-1}.$$

The early discoverers experienced serious difficulty with proving this formula. It is generally considered that the first complete proof was given around the year 1900 by Jules Henri Poincaré.

More relations among face numbers?

Having seen this formula it is natural to ask: *What other relations, if any, do the face numbers f_i satisfy*? This question opens the doors to a huge and very active research area, pursued by combinatorialists and geometers. Many equalities and inequalities are known for various classes of polytopes, such as upper bounds and lower bounds for the numbers f_i in terms of the dimension d and the number f_0 of vertices.

The boldest hope one can have for the study of f-vectors of polytopes is to obtain a complete characterization. By this is meant a reasonably simple set of conditions by which one can recognize if a given string of numbers is the f-vector of a d-polytope or not. For instance, one may ask whether

$$(14, 89, 338, 850, 1484, 1834, 1604, 971, 380, 76) \tag{12.1}$$

is the f-vector of a 10-dimensional polytope? We find that

$$14 - 89 + 338 - 850 + 1484 - 1834 + 1604 - 971 + 380 - 76 = 0,$$

in accordance with the generalized Euler formula. Had this failed we would know for sure that we are not dealing with a true f-vector, but agreeing with the Euler formula is certainly not enough to draw any conclusion. What other "tests" are there, strong enough to tell for sure whether this is the f-vector of a 10-polytope?

An answer is known for dimension 3; namely, (f_0, f_1, f_2) is the f-vector of a 3-polytope if and only if

$$\begin{aligned} &(i) && f_0 - f_1 + f_2 = 2, \\ &(ii) && f_0 \leq 2f_2 - 4, \\ &(iii) && f_2 \leq 2f_0 - 4. \end{aligned}$$

However, already the next case of 4 dimensions presents obstacles that with present methods are unsurmountable. Thus, no characterization of f-vectors of general polytopes is known. But if one narrows the class of polytopes to the so called "simplicial" ones there is a very substantial result that we now describe.

A *d-simplex* is a d-polytope which is cut out by exactly $d+1$ plane cuts. In other words, it has $d+1$ maximal faces, which is actually the minimum possible for a d-polytope. A 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, and so on; see Figure 34. In general, a d-simplex is the natural d-dimensional analogue of the tetrahedron.

A d-polytope is said to be *simplicial* if all its faces are simplices. It comes to the same to demand that all maximal faces are $(d-1)$-simplices. For instance, a 3-polytope is simplicial if all 2-faces are triangular, as in Figure 31; so the octahedron and icosahedron are examples of simplicial polytopes but the cube and permutohedron are not.

The class of simplicial polytopes is special from some points of view, but nevertheless very important in polytope theory. For instance, if one seeks to maximize the number of i-faces of a d-polytope with n vertices, the maximum is obtained simultaneously for all i by certain simplicial polytopes.

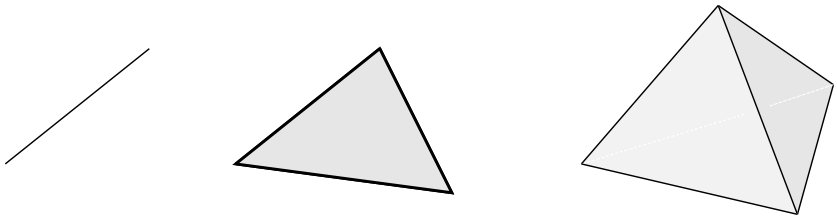


FIGURE 34
A d-simplex, $d = 1, 2, 3$.

In 1971 Peter McMullen made a bold conjecture for a characterization of the f-vectors of simplicial polytopes. A key role in his proposed conditions was played by certain "g-numbers," so his conjecture became known as the "g-conjecture." In 1980 two papers, one by Louis Joseph Billera and Carl William Lee and one by Richard Peter Stanley, provided the two major implications that were needed for a proof of the conjecture. Their combined efforts thus produced what is now known as the "g-theorem." To state the theorem we need to introduce an auxiliary concept.

By a *multicomplex* we mean a nonempty collection $\mathcal{M}$ of monomials in indeterminates $x_1, x_2, \ldots, x_n$ such that if $m \in \mathcal{M}$ and m' divides m then $m' \in \mathcal{M}$. Figure 35 shows the multicomplex $\mathcal{M} = \{1, x, y, z, x^2, xy, yz, z^2, x^2y, z^3\}$ ordered by divisibility.

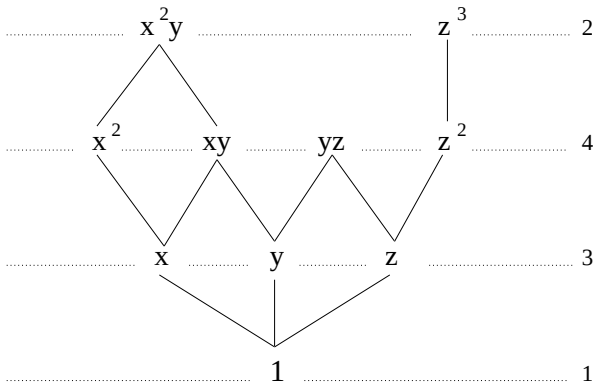


FIGURE 35
A multicomplex.

An *M-sequence* is a sequence $(1, a_1, a_2, a_3, \dots)$ such that each a_i is the number of monomials of degree i in some fixed multicomplex. For instance, the M-sequence coming from the multicomplex $\mathcal{M}$ in Figure 35 is $(1, 3, 4, 2)$. A multicomplex and an M-sequence can very well be infinite, but only finite ones will concern us here. If some zeros are added or removed at the end of a finite M-sequence it remains an M-sequence.

The "M" in M-sequence is mnemonic both for "multicomplex" and for "Macaulay," in honor of Francis Sowerby Macaulay, who studied the concept in a paper from 1927. Macaulay's purpose was entirely algebraic (to characterize so called Hilbert functions of certain graded algebras), but the underlying combinatorics of his investigations has turned out to have far-reaching ramifications.

THE g-THEOREM

We are now ready to formulate the theorem characterizing the f-vectors of simplicial d-polytopes. Let δ be the greatest integer less than or equal to $d/2$, and let $M_d = (m_{i,j})$ be the matrix with $(\delta + 1)$ rows and d columns and with entries

$$m_{i,j} = \binom{d+1-i}{d-j} - \binom{i}{d-j}, \quad \text{for } 0 \le i \le \delta,\ 0 \le j \le d-1.$$

Here we are once more using the binomial coefficients introduced in Chapter 1. Recall that they are given by

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!},$$

where $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$, and $0! = 1$.

For example, with $d = 10$ we get

$$m_{2,8} = \binom{10+1-2}{10-8} - \binom{2}{10-8} = \frac{9!}{2! \cdot 7!} - \frac{2!}{2! \cdot 0!} = 36 - 1 = 35,$$

and the whole matrix is

$$M_{10} = \begin{pmatrix} 11 & 55 & 165 & 330 & 462 & 462 & 330 & 165 & 55 & 11 \\ 1 & 10 & 45 & 120 & 210 & 252 & 210 & 120 & 45 & 9 \\ 0 & 1 & 9 & 36 & 84 & 126 & 126 & 84 & 35 & 7 \\ 0 & 0 & 1 & 8 & 28 & 56 & 70 & 55 & 25 & 5 \\ 0 & 0 & 0 & 1 & 7 & 21 & 34 & 31 & 15 & 3 \\ 0 & 0 & 0 & 0 & 1 & 5 & 10 & 10 & 5 & 1 \end{pmatrix}$$

These matrices M_d determine a very surprising link between M-sequences and f-vectors.

The g-theorem. *The matrix equation*

$$f = \boldsymbol{g} \cdot M_d$$

gives a one-to-one correspondence between f-vectors f of simplicial d-polytopes and M-sequences $\boldsymbol{g} = (g_0, g_1, \ldots, g_\delta)$.

The equation $f = \boldsymbol{g} \cdot M_d$ is to be understood as follows. Multiply each entry in the first row of M_d by g_0, then multiply each entry in the second row by g_1, and so on. Finally, after all these multiplications add the numbers in each column. Then the first column sum will equal f_0, the second column sum will equal f_1, and so on.

To exemplify the power of this theorem let us return to a question posed earlier; namely, is the vector f displayed in equation (12.1) the f-vector of a 10-polytope? This question can now be answered if sharpened from "10-polytope" to "simplicial 10-polytope". Easy computation shows that

$$f = (1, 3, 4, 2, 0, 0) \cdot M_{10}, \tag{12.2}$$

and we know from Figure 35 that $(1, 3, 4, 2, 0, 0)$ is an M-sequence. Hence, f is indeed the f-vector of some simplicial 10-polytope.

Having seen this, one can wonder if we were just lucky with this relatively small example. Perhaps for large d it is as hard to determine if a sequence is an M-sequence as to determine if a sequence is an f-vector coming from a simplicial polytope. This is not the case. There exists a very easy criterion in terms of binomial coefficients that quickly tests an integer sequence for being an M-sequence; see the appendix.

The proof of the g-theorem is very involved and calls on a lot of mathematical machinery. The part proved by Billera and Lee — that for every M-sequence $\boldsymbol{g}$ there exists a simplicial polytope with the corresponding f-vector — requires some very delicate geometrical arguments. The part proved by Stanley — that conversely to every simplicial polytope there corresponds an M-sequence in the stated way — uses tools from algebraic geometry in an essential way. Here is a brief statement for readers with sufficient background. There are certain complex projective varieties, called *toric varieties*, associated to d-polytopes with rational coordinates, and the fact that the sequence $\mathbf{g}$ corresponding to the f-vector of a polytope is an M-sequence ultimately derives from a multicomplex that can be constructed in the cohomology algebra of such a variety.

The g-vector associated to a simplicial polytope via the g-theorem is rich in geometric, algebraic and combinatorial meaning, yet it is still poorly understood and the subject of much current study.

In this book we have several times commented on the many surprising, remarkable and mysterious connections that exist between different mathematical objects, different mathematical problems and different mathematical areas. The g-theorem is one more example of this kind, establishing a totally unsuspected link between the combinatorial structure of multicomplexes of monomials and the facial structure of simplicial polytopes — two seemingly totally unrelated classes of objects.

In closing, let us once more mention that no characterization is known for f-vectors of general polytopes of dimension greater than 3. The success in the case of simplicial polytopes depends on some very special structure, available in that case but lacking or much more complex in general. The study of f-vectors, initiated by Euler's discovery some 250 years ago, is likely to remain an important challenge for many years to come.

APPENDIX

The characterization of M-sequences, referred to above, stems from an extremal question concerning the combinatorics of the divisibility relation for monomials.

Let $\mathcal{M}$ be a finite set of monomials of degree k in indeterminates $x_1, x_2, \ldots$. Define the *shadow* of $\mathcal{M}$, denoted $\partial(\mathcal{M})$, to be the set of monomials of degree $k-1$ that divide some element of $\mathcal{M}$. For instance, if

$$\mathcal{M} = \{x^3y, x^2y^2, x^2yz, xy^2z, z^3u, yzu^2\},$$

then

$$\partial(\mathcal{M}) = \{x^3, x^2y, xy^2, x^2z, xyz, y^2z, z^3, yzu, z^2u, yu^2, zu^2\}.$$

The question now is: *How small can the shadow be, given the size of* $\mathcal{M}$ *?* For example, the monomial family just shown has $\#\mathcal{M} = 6$ and $\#\partial(\mathcal{M}) = 11$, where $\#S$ denotes the number of elements of the set S. But choosing

$$\mathcal{M}' = \{x^4, x^3y, x^2y^2, xy^3, y^4, x^3z\},$$

we find that

$$\partial(\mathcal{M}') = \{x^3, x^2y, xy^2, y^3, x^2z\},$$

with $\#\mathcal{M}' = 6$ and $\#\partial(\mathcal{M}') = 5$. This is, in fact, best possible: it can be shown that $\#\partial(\mathcal{M}) \geq 5$ for all families $\mathcal{M}$ of 6 monomials of degree 4.

The general fact is that given the number of monomials $n = \#\mathcal{M}$ of a set $\mathcal{M}$ of monomials of degree k, there is an easily computable number $\partial^k(n)$ giving the minimal possible size of the shadow $\#\partial(\mathcal{M})$. Here is its definition.

For any integers $k, n \geq 1$ there is a unique way of writing

$$n = \binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \cdots + \binom{a_i}{i},$$

so that $a_k > a_{k-1} > \cdots > a_i \geq i \geq 1$. To see this, first choose a_k as large as possible so that $n \geq \binom{a_k}{k}$. Next, choose a_{k-1} as large as possible so that $n - \binom{a_k}{k} \geq \binom{a_{k-1}}{k-1}$, and continue like this to obtain the rest of the numbers a_j.

This given, define

$$\partial^k(n) = \binom{a_k - 1}{k-1} + \binom{a_{k-1} - 1}{k-2} + \cdots + \binom{a_i - 1}{i-1}.$$

Also let $\partial^k(0) = 0$.

The characterization we have referred to is this: A nonnegative integer sequence $(n_0, n_1, n_2, \ldots)$ is an *M-sequence* if and only if $n_0 = 1$ and

$$\partial^k(n_k) \leq n_{k-1} \qquad \text{for all } k > 1.$$

For example, we know from Figure 35 that $(1, 3, 4, 2, 0, 0)$ is an M-sequence. This fact can now be verified also by the the following small calculation:

$\partial^2(4) = 3 \leq 3$
$\partial^3(2) = 2 \leq 4$
$\partial^4(0) = 0 \leq 2$
$\partial^5(0) = 0 \leq 0.$

13
Connections with topology

On first acquaintance combinatorics may seem to have a somewhat different "flavor" than the mainstream areas of mathematics, due mainly to what mathematicians call "discreteness." Nevertheless, combinatorics is fortunate to have many beautiful and fruitful links with older and more established areas, such as algebra, geometry, probability and topology. We will now move on to discuss one such connection, perhaps the most surprising one, namely that with topology. First, however, let us say a few words about what mathematicians mean by discreteness.

In mathematics the words "continuous" and "discrete" have technical meanings that are quite opposite. Typical examples of continuous objects are curves and surfaces in ordinary 3-dimensional space (or suitable generalizations in higher dimensions). A characteristic property is that each point on such an object is surrounded by some "neighborhood" of other points, containing points that are in a suitable sense "near" to it. The area within mathematics that deals with the study of continuity is called *topology*. The characteristic property of discrete objects, on the other hand, is that each point is "isolated" — there is no concept of points being "near." Combinatorics is the area that deals with discreteness in its purest form, particularly in the study of finite structures of various kinds.

Several fascinating connections between the continuous and the discrete are

known in mathematics — in algebra, geometry and analysis. A quite recent development of this kind, the one we want to talk about here, is that ideas and results from topology can be put to use to solve certain combinatorial problems. In the following chapters we exemplify this with some problems coming from extremal combinatorics and computer science.

A nodding acquaintance with the most basic notions of metric topology, such as the meaning of "open set" , "continuous function" and "distance" between points in Euclidean space $\mathbf{R}^d$ will be assumed. The *d-dimensional sphere* $\mathbf{S}^d$ is the subspace of $\mathbf{R}^{d+1}$ consisting of points at distance one from the origin. In symbols: $\mathbf{S}^d = \{(x_1, x_2, \ldots, x_{d+1}) \in \mathbf{R}^{d+1} : x_1^2+x_2^2+\cdots+x_{d+1}^2 = 1\}$.

The Polish mathematician Karol Borsuk made some fundamental contributions to the early development of topology. In 1933 he published a paper entitled (in translation) "Three theorems about the n-dimensional Euclidean sphere". That paper contains, among other wonderful things, a famous theorem and a famous open problem. We need two definitions. Two points x, y on the sphere $\mathbf{S}^d$ are said to be *antipodal* if $x = -y$. For instance, the north pole and the south pole are antipodal on the surface of the earth, illustrating the $d = 2$ case. The *diameter* of a bounded set is the maximal distance between any pair of its points.

Borsuk's Theorem. [2]) *If the d-dimensional sphere $\mathbf{S}^d$ is covered by $d+1$ open sets, then one of these sets must contain a pair of antipodal points.*

Borsuk's Problem. *Is it true that every set of bounded diameter in d-dimensional real space $\mathbf{R}^d$ can be partitioned into $d+1$ sets of smaller diameter?*

This work of Borsuk has interacted with combinatorics in a remarkable way. In 1978 László Lovász solved a difficult combinatorial problem — the "Kneser Conjecture" from 1955 — using Borsuk's theorem. Then, in 1992 the debt to topology was repaid when Jeffry Ned Kahn and Gil Kalai solved Borsuk's problem using some results from pure combinatorics. Both of these achievements are discussed in chapter 14.

SIMPLICIAL COMPLEXES

We now outline in some detail another fruitful connection between topology and combinatorics, a connection that will be used in Chapters 15 and 16. Let

[2]) Also called the Borsuk-Lyusternik-Schnirelman theorem

us take as our example of a topological space the *torus*, a 2-dimensional surface that is well known in ordinary life in the form of an inner tube, or as the surface of a doughnut (see Figure 36).

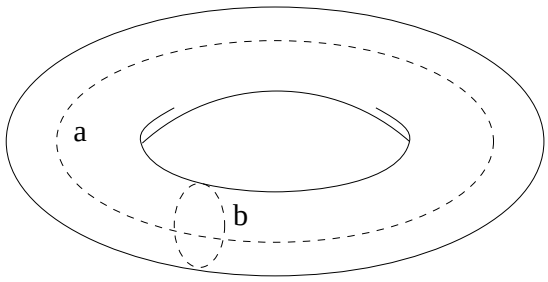


FIGURE 36
The torus

There is a way to "encode" a space such as the torus into a finite set system, called a *triangulation*. It works as follows. Draw (curvilinear) triangles on the torus so that each edge of a triangle is also the edge of some other triangle, and the 2 endpoints of each edge are not the pair of endpoints of any other edge. The triangles should cover the torus so that each point on the torus is in exactly one of the triangles, or possibly in an edge where two triangles meet or at a corner where several triangles meet. We can think of this as cutting the rubber surface of an inner tube into small triangular pieces. Figure 37 shows one way of doing this using 14 triangles. In this figure the torus is cut up and flattened out — to get back the original torus one has to roll this flattened version up and glue together the two sides marked 1-2-3-1, and then wrap around the cylinder obtained and glue together the two end-circles marked 1-4-5-1. Note that the two circles 1-2-3-1 and 1-4-5-1 in Figure 37 correspond to the circles marked **a** and **b** that are drawn with dashed lines on the torus in Figure 36.

Having thus cut the torus apart we now have a collection of 14 triangles. The corners in Figure 37 where triangles come together are called *vertices*, and we can represent each triangle by its 3 vertices. Thus each one of our 14 triangles is replaced by a 3-element subset of $\{1,2,3,4,5,6,7\}$. For instance, $\{1,2,4\}$ and $\{3,4,6\}$ denote two of the triangles. The full list of all 14 triangles is

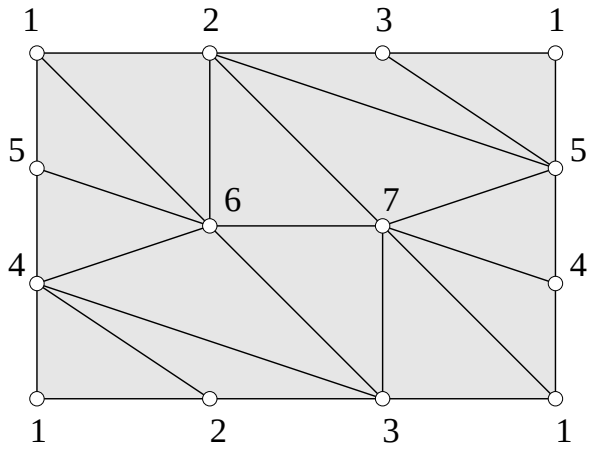


FIGURE 37
A triangulated torus

(13.1)
$$\begin{array}{lllllll} 124 & 126 & 135 & 137 & 147 & 156 & 234 \\ 235 & 257 & 267 & 346 & 367 & 456 & 457. \end{array}$$

A family of subsets of a finite set which is closed under taking subsets (i.e., if A is a set in the family and B is obtained by removing some elements from A then also B is in the family) is called a *simplicial complex*. Thus our fourteen 3-element sets and all their subsets form a simplicial complex. We encountered other examples in Chapter 12: if a polytope is simplicial then its faces form a simplicial complex.

An important fact is that just knowing the simplicial complex — a finite set system — we can fully reconstruct the torus! Namely, knowing the 14 triples we can manufacture 14 triangles with vertices marked in corresponding fashion and then glue these triangles together according to the blueprint of Figure 37 (using the vertex labels) to obtain the torus. To imagine this you should think of the triangles as being flexible (e.g., made of rubber sheet) so that there are no physical obstructions to their being bent and glued together. Also, the torus obtained may be different in size or shape from the original one (smaller, larger, deformed), but these differences are irrelevant from the point of view of topology.

To sum up the discussion: The simplicial complex coming from a triangulation is *a complete encoding* of the torus as a topological object. Every property of the torus that topology can have anything to say about is also a property of this finite set system!

BETTI NUMBERS

Why would topologists want to use such an encoding? The main reason is that they are interested in computing certain so-called *invariants* of topological spaces, such as the "Betti numbers" which we will soon comment on. The spaces they consider (such as the torus) are geometric objects with infinitely many points, on which it is usually hard to perform concrete computations. An associated simplicial complex, on the other hand, is a finite object which is easily adapted to computation (except possibly for size reasons). Topological invariants depend only on the space in question, but their computation may depend on choosing a triangulation or other "combinatorial decomposition". The part of topology that develops this connection is known as *combinatorial topology*. It was initiated by the great French mathematician Jules Henri Poincaré in the last years of the 1800's and greatly developed in the first half of the 20th century. Eventually the subject took on a more and more algebraic flavor, and in the 1940's the area changed name to *algebraic topology*.

The *Betti numbers* of a space are topological invariants that can be said to count the number of "independent holes" of various dimensions; the Betti number $\beta_j(T)$ is the number of $(j+1)$-dimensional such holes of the space T. It is impossible to give the full technical definition within the framework of this book. Let it suffice to say that the definition depends on certain algebraic constructions and to give some examples.

If T is a d-dimensional topological space then there are $d+1$ Betti numbers

$$\beta_0(T), \beta_1(T), ..., \beta_d(T),$$

which are nonnegative integers. Once we have a triangulation of a topological space the computation of Betti numbers is a matter of some (in principle) very simple linear algebra. (Note to specialists: Our $\beta_j(T)$'s are the *reduced* Betti numbers of T, differing from the "ordinary" Betti numbers only in that $\beta_0(T)+1$, rather than $\beta_0(T)$, is the number of connected components of T.)

For instance, the d-dimensional sphere has Betti numbers $(0, ..., 0, 1)$, reflecting the fact that it has exactly one $(d+1)$-dimensional "hole" (its interior) and no holes of other dimensions. The torus has Betti numbers $(0, 2, 1)$ because there are two essentially different 2-dimensional holes (spanned by

the circles **a** and **b** in Figure 36) and one 3-dimensional hole (the interior). Note that the two circles **a** and **b** delimit genuine "holes" in the sense that they cannot be continuously deformed to single points within the torus, and that they are "different" holes since one cannot be continuously deformed into the other.

We have seen that finite set systems are of use in topology as encodings of topological spaces. But the connection between topological spaces and simplicial complexes opens up a two-way street. What if the mathematics we are doing deals primarily with finite set systems, as is often the case in combinatorics? For instance, say that a combinatorial problem we are dealing with involves the fourteen 3-element sets listed in (13.1). Could the properties of the associated topological space — the torus — be of any relevance? For instance, could its Betti numbers (measuring the number of "holes" in the space) have something useful to say about the set system as such? We will exemplify that this may indeed be the case, and this is in fact one of the cornerstones for the "topological method" in combinatorics.

An example of a combinatorial concept which has topological meaning is the Möbius function of a partially ordered set, defined in Chapter 10. The connection is as follows. Let P be a poset with bottom element a and top element t, and let $\overline{P} = P \setminus \{a,t\}$, meaning P with a and t removed. Define the set family $\Delta(P)$ to consist of all *chains* (meaning: totally ordered subsets) $x_1 < x_2 < \cdots < x_k$ in $\overline{P}$. Then $\Delta(P)$ is a simplicial complex (since a subset of a chain is also a chain), so as discussed above there is an associated topological space.

For instance, let P be the divisor poset of the number 60 shown in Figure 20. Then $\overline{P} = P \setminus \{1,60\}$ has the following twelve maximal chains

$$
\begin{array}{c}
2 \text{ — } 4 \text{ — } 12 \\
2 \text{ — } 4 \text{ — } 20 \\
2 \text{ — } 6 \text{ — } 12 \\
2 \text{ — } 6 \text{ — } 30 \\
2 \text{ — } 10 \text{ — } 20 \\
2 \text{ — } 10 \text{ — } 30 \\
3 \text{ — } 6 \text{ — } 12 \\
3 \text{ — } 6 \text{ — } 30 \\
3 \text{ — } 15 \text{ — } 30 \\
5 \text{ — } 10 \text{ — } 20 \\
5 \text{ — } 10 \text{ — } 30 \\
5 \text{ — } 15 \text{ — } 30.
\end{array}
$$

These twelve triples of the simplicial complex should be thought of as describing twelve triangles that are to be glued together along common edges. This gives the topological space shown in Figure 38 — a 2-dimensional disc.

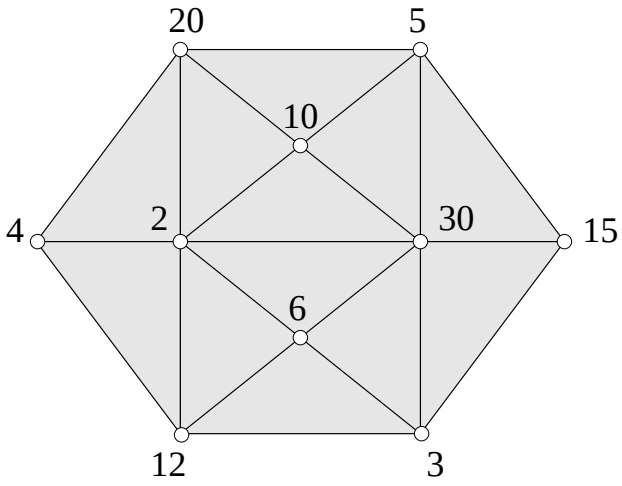


FIGURE 38
The simplicial complex of proper divisors of 60.

So, what does all this have to do with the Möbius function? The relation is this. Let $\beta_i(P)$ be the ith Betti number of the simplicial complex $\Delta(P)$, and let $\mu(P)$ denote the value that the Möbius function attains at the top element of P. Then,

$$\mu(P) = \beta_0(P) - \beta_1(P) + \beta_2(P) - \beta_3(P) + \cdots . \tag{13.2}$$

For instance, the space depicted in Figure 38 is a disc. The important thing here is that this space has no holes of any kind. Hence, all Betti numbers $\beta_i(P)$ are zero, implying via formula (13.2) that $\mu(P) = 0$. This "explains" topologically why $\mu(60) = 0$, a fact we already know from simpler considerations; see page 86. On the other hand, if P is the divisor diagram of the number 30 (which can be seen as a substructure in Figure 20), then $\Delta(P)$ is the circle 2 — 6 — 3 — 15 — 5 — 10 — 2 (a substructure in Figure 38). This circle has a one-dimensional hole, so $\beta_1(P) = 1$. All other Betti numbers are zero, hence formula (13.2) gives that $\mu(30) = -1$, another fact we already encountered on page 86.

14
Extremal combinatorics

Extremal problems arise in all areas of mathematics. The nature of an extremal problem, in its simplest and purest form, is that some quantity is kept fixed while another one is left free to vary, subject to certain given constraints. How large, or how small, can this second quantity be? For instance, take the following well-known extremal problem from continuous mathematics, the so-called *isoperimetric problem* for curves in the plane. This asks: what is the largest area enclosed by a simple closed curve of given fixed length ℓ? The answer is that the largest area is $\ell^2/4\pi$, and that this area is obtained if and only if the curve is a circle.

Extremal problems of various kinds permeate combinatorics and its applications. We already encountered an example in Chapter 12. The characterization of M-sequences, stated in the appendix to that chapter, stems from an extremal question concerning the combinatorics of the divisibility relation for monomials. Namely:

Given the size $n = \#\mathcal{M}$ of a set $\mathcal{M}$ of monomials of degree k, how small can its shadow $\partial(\mathcal{M})$ be?

The exact answer to this question is known, as explained on page 118.

Here is another example, this time concerning the combinatorics of the containment relation of finite sets. Recall from Chapter 13 that a family $\mathcal{F}$ of subsets of the set $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$ is called a *simplicial complex* if it is

closed under taking subsets, that is, if $A \in \mathcal{F}$ and $B \subseteq A$ implies that $B \in \mathcal{F}$. Such set families were discussed there for their topological significance. The intersection of two simplicial complexes $\mathcal{F}$ and $\mathcal{G}$ is again a simplicial complex. This prompts the purely combinatorial extremal question:

Given the sizes $f = \#\mathcal{F}$ and $g = \#\mathcal{G}$ of two simplicial complexes $\mathcal{F}$ and $\mathcal{G}$ consisting of subsets of $[n]$, how small can their intersection $\mathcal{F} \cap \mathcal{G}$ be?

The following answer[3]) was given in 1966 by Daniel J. Kleitman:

$$\#(\mathcal{F} \cap \mathcal{G}) \geq \frac{gf}{2^n}. \tag{14.1}$$

The area of extremal combinatorics abounds with open problems, often easy to state but difficult to answer. For an example of an extremal-type question whose answer is not known, consider this deceptively simple-sounding conjecture proposed by Péter Frankl in 1979:

If a finite family $\mathcal{F}$ of nonempty sets is closed under taking unions, then some element belongs to at least half of the sets in $\mathcal{F}$.

Frankl's conjecture has been verified for small set families, more precisely if the size of the largest set in the family is at most 11 or if there are at most 36 sets in the family. It remains open in general.

Intersecting set families

Among many interesting topics in extremal combinatorics, we choose to focus on a few results concerning a very general class of set systems. We consider subsets of a given finite set, which for definiteness can be taken to be the set $[n] = \{1, 2, \ldots, n\}$. The family of *all* subsets of $[n]$ is denoted by $2^{[n]}$. A family $\mathcal{F}$ of subsets of $[n]$ is said to be *intersecting* if $A \cap B \neq \emptyset$ for all sets $A \in \mathcal{F}$ and $B \in \mathcal{F}$. The basic extremal question is:

How large can an intersecting family be?

The path to answering this question begins with the observation that the family $\mathcal{F}(i)$ of all subsets containing a fixed element $i \in [n]$ is intersecting, since $i \in A \cap B$ for all $A, B \in \mathcal{F}(i)$. The number of sets in $\mathcal{F}(i)$ is 2^{n-1}. Can a larger intersecting family be found? An application of the pigeonhole

[3]) Readers familiar with probability theory may observe that if written

$$\frac{\#(\mathcal{F} \cap \mathcal{G})}{2^n} \geq \frac{\#\mathcal{F}}{2^n}\frac{\#\mathcal{G}}{2^n}$$

this inequality has a probabilistic interpretation, namely it shows that the two events that a random subset of $[n]$ belongs to $\mathcal{F}$ or to $\mathcal{G}$ are positively correlated.

principle (stated on page 46) shows that the answer is no. Here is how: the set of all subsets of $[n]$ is naturally partitioned into disjoint pairs $\{A, A^c\}$, consisting of a subset A and its complement $A^c = [n] \setminus A$. Since there are 2^n subsets of $[n]$ there are 2^{n-1} such pairs. Thus, a set family $\mathcal{F}$ with more than 2^{n-1} members must contain both elements A and A^c of some pair. But, since $A \cap A^c = \emptyset$ this means that $\mathcal{F}$ is not intersecting.

We have proved the following simple result: *Let $\mathcal{F}$ be an intersecting family of subsets of $[n]$. Then $\#\mathcal{F} \leq 2^{n-1}$, and this upper bound is best possible.*

Let us now be a bit more ambitious and ask:

How large can the union of j intersecting set families be?

A moments thought suggests that the set family $\mathcal{F}(1) \cup \cdots \cup \mathcal{F}(j)$ is a possible candidate for being an extremal configuration. What is its size? For a set not to belong to $\mathcal{F}(1) \cup \cdots \cup \mathcal{F}(j)$ is the same as to be a subset of $\{j+1, j+2, \ldots, n\}$, and there are 2^{n-j} such subsets. Hence,

$$\#(\mathcal{F}(1) \cup \cdots \cup \mathcal{F}(j)) = 2^n - 2^{n-j}.$$

The following result shows that these set families are indeed extremal.

Kleitman's Theorem. *Let $\mathcal{F}$ be the union of j intersecting families of subsets of $[n]$. Then*

$$\#\mathcal{F} \leq 2^n - 2^{n-j},$$

and this upper bound is best possible.

The $j = 1$ case, that is, the upper bound $\#\mathcal{F} \leq 2^{n-1}$ for intersecting families, was easy to prove using nothing but the pigeonhole principle. We now sketch a proof for the $j = 2$ case. This proof contains all the new ideas needed to deal with the general case.

Assume that $\mathcal{F}$ is the union of two intersecting set families $\mathcal{F}_1$ and $\mathcal{F}_2$. Enlarge $\mathcal{F}_1$ to a set family $\mathcal{G}_1$ by adding all subsets of $[n]$ that contain some set from $\mathcal{F}_1$:

$$\mathcal{G}_1 \stackrel{\text{def}}{=} \{A \in 2^{[n]} \mid A \supseteq B \text{ for some } B \in \mathcal{F}_1\}.$$

The enlarged set family $\mathcal{G}_1$ is also intersecting, so we know that $\#\mathcal{G}_1 \leq 2^{n-1}$. Hence, the complementary set family $\overline{\mathcal{G}_1} \stackrel{\text{def}}{=} 2^{[n]} \setminus \mathcal{G}_1$ satisfies

$$\#\overline{\mathcal{G}_1} \geq 2^n - 2^{n-1} = 2^{n-1}.$$

Construct $\mathcal{G}_2$ from $\mathcal{F}_2$ in the same way. We have from elementary set theory that

$$\#(\mathcal{F}_1 \cup \mathcal{F}_2) \le \#(\mathcal{G}_1 \cup \mathcal{G}_2) = 2^n - \#(\overline{\mathcal{G}_1} \cap \overline{\mathcal{G}_2}).$$

The crucial observation now is that $\overline{\mathcal{G}_1}$ and $\overline{\mathcal{G}_2}$ are closed under taking subsets, since clearly their complements $\mathcal{G}_1$ and $\mathcal{G}_2$ are closed under taking supersets. Hence, using equation (14.1) we get

$$\#(\overline{\mathcal{G}_1} \cap \overline{\mathcal{G}_2}) \ge \frac{\#\overline{\mathcal{G}_1} \cdot \#\overline{\mathcal{G}_2}}{2^n} \ge \frac{2^{n-1} \cdot 2^{n-1}}{2^n} = 2^{n-2},$$

and so,

$$\#\mathcal{F} = \#(\mathcal{F}_1 \cup \mathcal{F}_2) \le 2^n - \#(\overline{\mathcal{G}_1} \cap \overline{\mathcal{G}_2}) \le 2^n - 2^{n-2},$$

completing the proof of the $j = 2$ case.

In the study of intersecting set families it is natural and fruitful to restrict the "universe" of subsets considered for our extremal questions to families of sets of some fixed size k (rather than of all sizes). The most basic question then is the following.

How large can an intersecting family of k-element subsets be ?

Let $\binom{[n]}{k}$ denote the family of all k-element subsets of the set $[n] = \{1, 2, \ldots, n\}$. If $k > n/2$ then $\binom{[n]}{k}$ is itself an intersecting family, so to get nontrivial questions we must demand that $k \le n/2$. Guided by our previous reasoning it is natural to guess that $\mathcal{F}_k(i) \stackrel{\text{def}}{=} \mathcal{F}(i) \cap \binom{[n]}{k}$ —that is, the family of k-element subsets containing some fixed element i — is extremal. This turns out indeed to be the case, as was shown by Pál Erdős, Chao Ko, and Richard Rado in 1961. Note that

$$\#\mathcal{F}_k(i) = \binom{n-1}{k-1},$$

since a k-element subset of $[n]$ containing i is obtained by choosing $k-1$ elements among the $n-1$ elements other than i.

Erdős-Ko-Rado Theorem. *Let $\mathcal{F}$ be an intersecting family of k-element subsets of $[n]$, with $k \le n/2$. Then*

$$\#\mathcal{F} \le \binom{n-1}{k-1},$$

and this upper bound is best possible.

We leave aside a discussion of how to prove this basic result. Instead, with the case of a single family settled, we can in the same spirit that led to Kleitman's theorem move on to the more ambitious question:

How large can the union of j intersecting families of k-element subsets be ?

Reasoning along familiar lines we are led to guess that $\mathcal{F}_k(1) \cup \cdots \cup \mathcal{F}_k(j)$ might be an extremal set family. Since

$$\#\left(\mathcal{F}_k(1) \cup \cdots \cup \mathcal{F}_k(j)\right) = \binom{n}{k} - \binom{n-j}{k},$$

a plausible conjecture would therefore be that the union $\mathcal{F}$ of j intersecting families of k-element subsets must satisfy

$$\#\mathcal{F} \leq \binom{n}{k} - \binom{n-j}{k}.$$

This was shown to be true by Pál Erdős in 1965, but only for large enough n. It is unfortunately false in general, and with questions of this type we enter largely unknown territory. As an example of something that *is* known, we have that the union of $n - 2k + 1$ or fewer intersecting families of k-element subsets has size strictly smaller than $\binom{n}{k}$. This special case is the Lovász-Kneser theorem, to be discussed in the next section.

There are many variations possible on the theme of asking extremal questions for set families with constrained intersection properties. For instance, instead of forbidding intersections of size zero (as is done for intersecting families), one can forbid intersections of some other size. As a final glimpse of this branch of extremal combinatorics, we quote a 1981 theorem of Péter Frankl and Richard Michael Wilson, which has this flavor.

Frankl-Wilson Theorem. *Let q be a power of a prime number, and let $\mathcal{F}$ be a family of $2q$-element subsets of a $4q$-element set such that no two members of $\mathcal{F}$ have intersection of size q. Then,*

$$\#\mathcal{F} \leq 2 \cdot \binom{4q-1}{q-1}.$$

KNESER'S PROBLEM

In 1955 Martin Kneser published a tantalizing extremal problem concerning intersecting families. It is based on the observation that the family of all k-element subsets of an n-element set can be obtained as the union of $n-2k+2$ intersecting subfamilies. Namely, consider the following union:

$$\binom{[n]}{k} = \binom{[2k-1]}{k} \cup \mathcal{F}_k(2k) \cup \cdots \cup \mathcal{F}_k(n). \tag{14.2}$$

The greatest element m of a k-element subset of the set $[n]$ satisfies either $m < 2k$, in which case the set belongs to the first subfamily of the union, or $m \geq 2k$, in which case the set belongs to one of the other subfamilies. Kneser's conjecture (as it came to be known) is that fewer subfamilies are not possible for such a union. This was proved by László Lovász in 1978.

Lovász-Kneser Theorem. *If the family of all k-element subsets of an n-element set, with $k \leq n/2$, is obtained as the union of intersecting subfamilies, then at least $n - 2k + 2$ such subfamilies are needed.*

Why can we not do better? An interesting feature of Lovász' proof, and of other subsequent proofs of Kneser's conjecture, is that they rely on methods from topology. By sketching the relevant arguments and ideas we want to give a small glimpse of these interactions, which are quite unexpected.

An elegant and instructive way to view the problem is in terms of graph coloring. Define the *Kneser graph* $KG(n, k)$ to be the graph whose vertices are the k-element subsets of $[n]$ and whose edges are the pairs of disjoint such subsets. The graph $KG(5, 2)$ is shown in Figure 39.

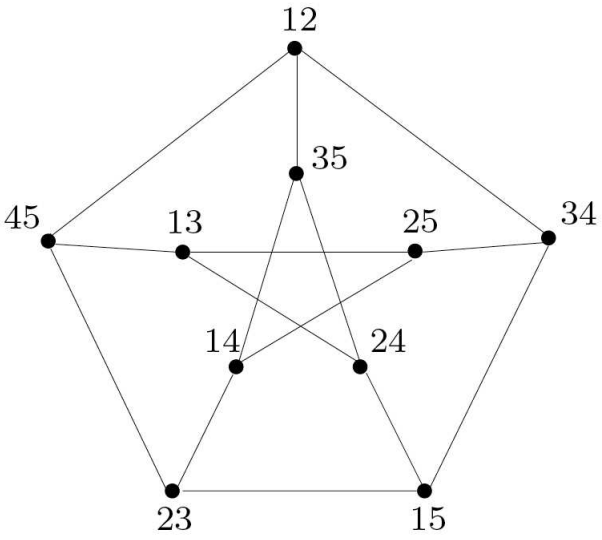


FIGURE 39
The Kneser graph $KG(5, 2)$.

The union (14.2) shows that the graph $KG(n,k)$ can be properly colored with $n-2k+2$ colors, meaning that one of $n-2k+2$ colors can be assigned to each vertex in such a manner that vertices joined by an edge receive distinct colors. For instance, the graph $KG(5,2)$ can be properly colored with 3 colors, as induced by the union (14.2):

$$\binom{[5]}{2} = \binom{[3]}{2} \cup \mathcal{F}_2(4) \cup \mathcal{F}_2(5). \tag{14.3}$$

A moment's reflection shows that the requirement for being a proper coloring is precisely that each color class forms an intersecting family. Thus we get the following reformulation of the theorem.

Lovász-Kneser Theorem (second version). *The Kneser graph $KG(n,k)$ cannot be properly colored with $n-2k+1$ colors.*

Here is how this conclusion is derived from Borsuk's theorem (stated on page 120). This version of the proof is due to Imre Bárány.

Assume first that $n = 2k+1$, and as the points in our ground set, take n points evenly spaced on a circle (see Fig. 40). Call this set X.

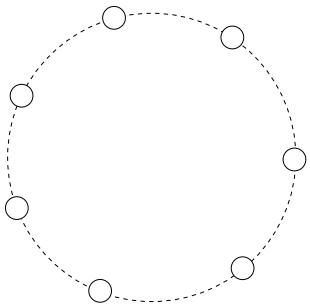


FIGURE 40
The set X for $n = 7$, $k = 3$.

Assume that to each k-element subset of X we have assigned either the color red or the color blue. Let R be the set of points r on the circle such that the open half-circle centered at r contains a red k-subset of X, and let B be similarly defined for the color blue.

Because of the even distribution of the points in X, it is clear that every point on the circle belongs either to R or to B (or to both). In other words,

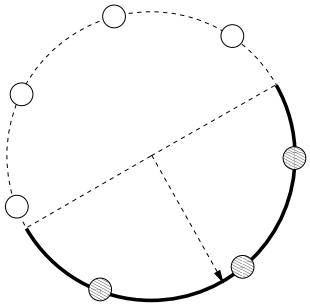


FIGURE 41
Open half-circle centered at the arrow point.

the sets R and B cover the circle. Also, one easily sees that the sets R and B are open.

Thus, since the circle is the same thing as the one-dimensional sphere $\mathbf{S}^1$, Borsuk's theorem tells us that there must exist a pair of antipodal points of the same color. If, say both x and $-x$ belong to R, this means that the two half-circles centered at these points each contains a "red" k-set. But these half-circles are disjoint. Therefore the two red k-sets are also disjoint. The existence of a pair of disjoint k-subsets of the same color is precisely what we wanted to prove, so this settles the $n = 2k + 1$ case.

The general proof for $n > 2k + 1$ is entirely similar, the only difference being that the argument takes place in higher dimensions. Namely, one chooses a subset X of the $(n - 2k)$-dimensional sphere and assigns one of $n - 2k + 1$ colors to each k-element subset of X. Some care has to be taken in order to secure that the points of X are, in a suitable sense, evenly spaced. Then the argument goes through along the same lines, now using Borsuk's theorem for $\mathbf{S}^{n-2k}$.

BORSUK'S PROBLEM

The Frankl-Wilson theorem (page 131) has an unsuspected application. Recall Borsuk's problem (page 120):

> *Is it true that every set of bounded diameter in k-dimensional Euclidean space $\mathbf{R}^k$ can be partitioned into $k + 1$ sets of smaller diameter?*

The answer is definitely "yes" when $k = 1$; the statement then comes down to dividing a line segment of length 1 into two shorter segments. It was also long known that the statement is true for $k = 2$ and $k = 3$, and it

was generally believed that the statement is true for all dimensions k — this became known as *Borsuk's conjecture*.

It therefore came as a great surprise when in 1992 Jeffry Ned Kahn and Gil Kalai showed that the answer to Borsuk's problem is actually "no," contrary to what "everyone" had believed for nearly 60 years. The key to solving Borsuk's geometric problem comes from combinatorics, more precisely, it is provided by the Frankl-Wilson theorem. Here is the argument of Kahn and Kalai.

The ground set for the construction is the set E of all edges in the complete graph on vertex set $[4q]$, where q is a power of a prime number. Thus,

$$\#E = \binom{4q}{2} = 2q(4q-1).$$

Let P be the set of all partitions of $[4q]$ into two disjoint subsets A and B of equal size $2q$. We denote by $\langle A, B\rangle$ such an element of P. Note that $\langle A, B\rangle = \langle B, A\rangle$ and that

$$\#P = \frac{1}{2}\binom{4q}{2q}.$$

For each such partition $\langle A, B\rangle$, let $E\langle A, B\rangle$ be the subset of E consisting of the edges with one endpoint in A and the other in B. Then $\#(E\langle A, B\rangle) = 4q^2$, since each endpoint of such an edge can be independently chosen in $2q$ different ways.

For $\langle A, B\rangle \in P$ let $x^{\langle A,B\rangle} \in \mathbf{R}^E$ be the point defined coordinate-wise by

$$x_i^{\langle A,B\rangle} = \begin{cases} 1, & \text{if } i \in E\langle A, B\rangle \\ 0, & \text{if } i \notin E\langle A, B\rangle \end{cases}$$

and let $X = \{x^{\langle A,B\rangle} : \langle A, B\rangle \in P\}$. A small computation, which we defer to the appendix, shows that the diameter of the set X is $2q$, and that the distance between two points $x^{\langle A,B\rangle}$ and $x^{\langle C,D\rangle}$ in X equals $2q$ if and only if $\#(A \cap C) = q$.

The idea of the proof is now the following. Suppose that the set X is divided into r pieces, where

$$r < \frac{\#X}{2\binom{4q-1}{q-1}}$$

Then one piece will contain more than $2\binom{4q-1}{q-1}$ points $x^{\langle A,B\rangle}$, so by the Frankl-Wilson theorem there must among them be two points $x^{\langle A,B\rangle}$ and $x^{\langle C,D\rangle}$ such that $\#(A \cap C) = q$. In other words, one piece of the partition of the set X into r pieces must have the same diameter $2q$ as X itself.

Thus, the construction yields a set $X \in \mathbf{R}^E$ violating Borsuk's conjecture if

$$\dim \mathbf{R}^E + 1 \;<\; \frac{\#X}{2\binom{4q-1}{q-1}},$$

that is,

$$2q(4q-1)+1 \;<\; \frac{\frac{1}{2}\binom{4q}{2q}}{2\binom{4q-1}{q-1}}.$$

This inequality can be shown to be satisfied for q large enough, yielding counterexamples to Borsuk's conjecture in $2,014$-dimensional space.

Later research has reduced the dimension of counterexamples from $2,014$ to 298. Borsuk's conjecture is, however, still undecided in dimension 4.

APPENDIX

The distance between points in the set X is combinatorially determined in the following way. Since the distance $\operatorname{dist}(x,y)$ between two points x and y in $\mathbf{R}^n$ is

$$\operatorname{dist}(x,y) = \sqrt{(x_1-y_1)^2+\cdots+(x_n-y_n)^2},$$

we have that

$$\begin{aligned}
\operatorname{dist}(x^{\langle A,B\rangle}, x^{\langle C,D\rangle})^2 &= \#(E\langle A,B\rangle \cup E\langle C,D\rangle) - \#(E\langle A,B\rangle \cap E\langle C,D\rangle) \\
&= \#E\langle A,B\rangle + \#E\langle C,D\rangle - 2\cdot\#(E\langle A,B\rangle \cap E\langle C,D\rangle) \\
&= 2\cdot 4q^2 - 2(\#(A\cap C)\cdot\#(B\cap D) + \#(A\cap D)\cdot\#(B\cap C)).
\end{aligned}$$

Writing $z = \#(A\cap C)$ (for ease of notation) and observing that

$$\begin{aligned}
\#(A\cap C) = \#(B\cap D) &= z \\
\#(A\cap D) = \#(B\cap C) &= 2q - z,
\end{aligned}$$

we get

$$\operatorname{dist}(x^{\langle A,B\rangle}, x^{\langle C,D\rangle})^2 = 8q^2 - 2(z^2+(2q-z)^2) = 4z(2q-z) \le 4q^2.$$

Thus $\operatorname{dist}(x^{\langle A,B\rangle}, x^{\langle C,D\rangle}) \le 2q$ with equality if and only if $\#(A\cap C) = q$.

15
Complexity of sorting and distinctness

A major theme in theoretical computer science is to estimate the complexity of computational tasks. By "complexity" is here meant the amount of time and of computational resources needed. To show that a task can be done in a certain number of steps one must construct an algorithm achieving the task and prove that it performs as claimed. It is often the more difficult part of the problem to show that there is no "faster" way, i.e., requiring fewer steps. Examples of this are given in this chapter and the next.

How many comparisons needed?

The following is a very basic situation studied in complexity theory. A sequence of real numbers $x_1, x_2, \ldots, x_n$ is given. A computer is asked to decide some property of the sequence or to restructure it using only pairwise comparisons. This means that the computer is allowed to learn about the input sequence only by inspecting pairs x_i and x_j and deciding whether $x_i > x_j$, $x_i < x_j$ or $x_i = x_j$. The question then is: How many such comparisons must the computer perform in the worst case when using the best algorithm? This number, as a function of n, is called the *complexity* of the problem.

The following notation is used to state such results. To say that the complexity is $\Theta(f(n))$, where $f(n)$ is some function, means that there exist constants c_1 and c_2 such that

$$c_1 \cdot f(n) < \text{complexity} < c_2 \cdot f(n).$$

While this notation doesn't give the exact numerical value of the complexity (which is often hard, if not impossible, to determine) it reveals its order of growth, which is what is usually taken as the main indication if a problem is computationally easy or hard. In the following formulas the function "$\log n$" will frequently appear. Readers not familiar with the logarithm function can take this to mean roughly the number of digits needed to write the number n in base 10, so that for instance $\log 1997 \approx 4$.

Here are some basic and well-known examples.

1. **Sorting.** *To rearrange the n numbers increasingly $x_{i_1} \le x_{i_2} \le \cdots \le x_{i_n}$ requires $\Theta(n \log n)$ comparisons.*
2. **Median.** *To find j such that x_j is "in the middle," meaning that half of the x_i's are less than or equal to x_j and half of the x_i's are greater than or equal to x_j, requires $\Theta(n)$ comparisons. In fact, it has been shown that at least $2n$ comparisons are needed and that at most $3n$ comparisons suffice.*
3. **Distinctness.** *To decide whether all entries x_i are distinct, that is whether $x_i \neq x_j$ when $i \neq j$, requires $\Theta(n \log n)$ comparisons.*

The problem we wish to discuss is a generalization of the distinctness problem. Namely,

k-equal problem: *for $k \ge 2$, decide whether some k entries are equal, that is, can we find $i_1 < i_2 < \cdots < i_k$ such that $x_{i_1} = x_{i_2} = \cdots = x_{i_k}$?*

For example, are there nine equal entries in the following list of one-digit numbers?

24791374685848713955196742346159463314867729559243628541178369725819 32

Answer: Yes, there are nine copies of the number "4". Are there ten equal entries? Answer: No. If pairwise comparisons are the only type of operation allowed, how should one go about settling these questions in an efficient manner, and how many comparisons would be needed?

Here are a few immediate observations. If $k = 2$ the problem reduces to the distinctness problem, so the complexity is $\Theta(n \log n)$. At the other end of the scale, if $k > \frac{n}{2}$ the complexity is $\Theta(n)$, because we can argue as follows. The median x_j can be found using $3n$ comparisons. If there are $k > \frac{n}{2}$ equal entries then the median must be one of them. Thus after comparing x_j with the other $n-1$ entries x_i we gain enough information to conclude whether there are some k entries that are equal. This procedure requires in all $4n-1$ comparisons. On the other hand it is easy to see that at least $n-1$ comparisons

are needed in the worst case, so there are both upper and lower bounds of the form "constant times n" to the complexity.

We have seen that the complexity of the k-equal problem decreases from $\Theta(n \log n)$ to $\Theta(n)$ when the parameter k grows from 2 to above $\frac{n}{2}$, so the k-equal problem seems to get easier the larger k gets. The exact form of this relationship is given in the following result from 1992 of Anders Björner, László Lovász, and Andrew Chi-Chih Yao.

Theorem. *The complexity of the k-equal problem is* $\Theta(n \log \frac{2n}{k})$.

The upper bound is obtained via a partial sorting algorithm based on repeated median-finding. It generalizes what was described for the case $k > \frac{n}{2}$ above. We shall leave it aside.

PROOF OF THE LOWER BOUND

The lower bound — proving that at least $n \log \frac{2n}{k}$ comparisons are needed (up to some constant) by *every* algorithm in the worst case — is the difficult and mathematically more interesting part. The proof uses a combination of topology and combinatorics. A detailed description would take us too far afield, but we will attempt to get some of the main ideas across.

Let us look at the situation from a geometric point of view. Each equation $x_{i_1} = x_{i_2} = \cdots = x_{i_k}$ determines an $(n-k+1)$-dimensional linear subspace of $\mathbf{R}^n$, the n-dimensional space consisting of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers x_i. The k-equal problem is from this point of view to determine whether a given point $x = (x_1, x_2, \ldots, x_n)$ lies in at least one such subspace, or — which is the same — lies in the union of all the subspaces $x_{i_1} = x_{i_2} = \cdots = x_{i_k}$.

Removal of linear subspaces disconnects $\mathbf{R}^n$. For instance, removal of a plane (a 2-dimensional subspace) cuts $\mathbf{R}^3$ into two pieces, whereas removal of a line (a 1-dimensional subspace) leaves another kind of "hole". These are precisely the kinds of holes that are measured by the topological Betti numbers (as was discussed in Chapter 13). Going back to the general situation, it seems clear that if *all* the subspaces $x_{i_1} = x_{i_2} = \cdots = x_{i_k}$ are removed from $\mathbf{R}^n$ then lots of holes of different dimensions will be created. This must mean that the sum of Betti numbers of $M_{n,k}$, the part of space $\mathbf{R}^n$ that remains after all these subspaces have been removed, is a large number:

$$\beta(M_{n,k}) = \beta_0(M_{n,k}) + \beta_1(M_{n,k}) + \cdots + \beta_n(M_{n,k}).$$

The idea now is that if the space $M_{n,k}$ is complicated topologically, as measured by this sum of Betti numbers, then this ought to imply that it is

computationally difficult to determine whether a point x belongs to it. This turns out to be true in the following quantitative form.

Fact 1. *The complexity of the k-equal problem is at least* $\log_3 \beta(M_{n,k})$.

Here $\log_3$ denotes logarithm to the base 3, which differs by a constant factor from the logarithm to the base 10 that was mentioned earlier.

So, now the problem has been converted into a topological one — to compute or estimate the sum of Betti numbers $\beta(M_{n,k})$. This can be done via a formula of Robert Mark Goresky and Robert Duncan MacPherson, which expresses these Betti numbers in terms of some finite simplicial complexes associated to partitions. To get further we need to introduce a few more concepts from combinatorics.

In Chapter 10 we discussed partitions of sets, and we shall return once more to the ubiquitous concept of partitions.

In the following we use $[n] = \{1, 2, \ldots, n\}$ as the ground set and for fixed k (an integer between 2 and n) consider the collection of all partitions of $[n]$ that have no parts of sizes $2, 3, \ldots, k-1$. Denote this collection by $\Pi_{n,k}$. For instance, $\Pi_{4,2}$ is the collection of *all* partitions of $\{1, 2, 3, 4\}$ (there are no forbidden parts), while $\Pi_{4,3}$ is the following subcollection (now parts of size 2 are forbidden):

1234, 1-234, 2-134, 3-124, 4-123, 1-2-3-4

Recall from Chapter 10 that there is a natural way to compare set partitions (a partition π is less than partition σ if π is obtained from σ by further partitioning its parts). This way we get an order structure on the set $\Pi_{n,k}$, which can be illustrated in a diagram. Figure 21 shows the order diagram of $\Pi_{4,2}$ and Figure 42 shows that of $\Pi_{4,3}$.

Now, consider the Möbius function (defined in Chapter 10) computed over the poset $\Pi_{n,k}$. Let $\mu_{n,k}$ denote the value attained by the Möbius function at the partition with only one part, which is at the top of the order diagram. For example, direct computation (as demonstrated in Chapter 10) over the posets in Figures 21 and 42 shows that $\mu_{4,3} = 3$ and $\mu_{4,2} = -6$.

We can now return to the discussion of the k-equal problem. We left off with the question of how to estimate the sum of Betti numbers $\beta(M_{n,k})$. The formula of Goresky and MacPherson mentioned earlier implies, by an argument involving among other things the topological significance of the Möbius function (discussed in Chapter 13), the following relation:

Fact 2. $\beta(M_{n,k}) \geq |\mu_{n,k}|$.

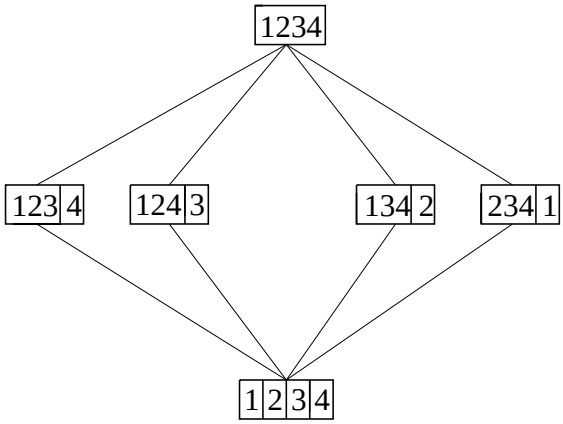


FIGURE 42
Order diagram of $\Pi_{4,3}$

Putting Facts 1 and 2 together, the complexity question for the k-equal problem has been reduced to the problem of showing that the combinatorially defined Möbius numbers $|\mu_{n,k}|$ grow sufficiently fast. For this we turn to the method of generating functions, already introduced in the early chapters on counting number partitions. Certain recurrences for the numbers $\mu_{n,k}$ lead, when interpreted at the level of generating functions, to the following formula:

$$\exp\left(\sum_{n\geq 1}\mu_{n,k}\frac{x^n}{n!}\right) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{k-1}}{(k-1)!}. \tag{15.1}$$

To make sense of this you have to imagine inserting the series $y = \sum_{n\geq 1}\mu_{n,k}\frac{x^n}{n!}$ into the exponential series $\exp(y) = \sum_{n\geq 0}\frac{y^n}{n!}$, and then expanding in powers of x. Also, since $\mu_{n,k}$ has so far been defined only for $k \leq n$ we should mention that we put $\mu_{n,k} = 0$ for $1 < n < k$ and $\mu_{1,k} = 1$.

From this relation between the numbers $\mu_{n,k}$ and the polynomial on the right-hand-side (which is a truncation of the exponential series) we can extract the following explicit information.

Fact 3. *Let* $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$ *be the complex roots of the polynomial* $1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{k-1}}{(k-1)!}$. *Then*

$$\mu_{n,k} = -(n-1)!\left(\alpha_1^{-n} + \alpha_2^{-n} + \cdots + \alpha_{k-1}^{-n}\right).$$

For instance, if $k = 2$ there is only one root $\alpha_1 = -1$, and we get

$$\mu_{n,2} = (-1)^{n-1}(n-1)!$$

— a fact we already encountered in equation (10.1). Also, in this case formula (15.1) specializes to

$$\exp\left(\sum_{n\geq 1}(-1)^{n-1}\frac{x^n}{n}\right) = 1 + x,$$

which is well-known to all students of the calculus in the equivalent form

$$\log(1+x) = \sum_{n\geq 1}(-1)^{n-1}\frac{x^n}{n}.$$

If $k = 3$ there are 2 roots $\alpha_1 = -1 + i$ and $\alpha_2 = -1 - i$, where $i = \sqrt{-1}$, and using some formulas from elementary complex algebra we get

$$\begin{aligned} \mu_{n,3} &= -(n-1)!\left((-1+i)^{-n} + (-1-i)^{-n}\right) \\ &= -(n-1)!\, 2^{1-\frac{n}{2}} \cos\frac{3\pi n}{4}. \end{aligned} \tag{15.2}$$

We have come to a point where we know on the one hand from Facts 1 and 2 that

$$\textit{the complexity of the k-equal problem} \geq \log_3 |\mu_{n,k}|,$$

and on the other that the Möbius numbers $\mu_{n,k}$ are given in terms of the roots $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$ as stated in Fact 3. It still remains to show that the numbers $|\mu_{n,k}|$ are large enough so that $\log_3 |\mu_{n,k}|$ produces the desired complexity lower bound. For this reason it comes as a chilling surprise to discover that these numbers are not always very large. In fact, formula (15.2) shows that

$$\mu_{n,3} = 0, \quad \text{for } n = 6, 10, 14, 18, 22, \ldots.$$

It can also be shown that $\mu_{2k,k} = 0$ for all odd numbers k.

So, we are not quite done — but almost! With a little more work it can be shown from the facts presented so far that $|\mu_{n,k}|$ is, so to say, "sufficiently large for sufficiently many n" (for fixed k). With this, and a "monotonicity argument" to handle the cases where $|\mu_{n,k}|$ itself is not large but nearby values are, it is possible to wrap up the whole story and obtain the initially stated lower bound of the form "constant times $n \log \frac{2n}{k}$".

Let us mention in closing that it is possible to work with Betti numbers the whole way, never passing to the Möbius function as described here. This route is a bit more complicated but results in a better constant for the lower bound.

16
Complexity of graph properties

In this final chapter we consider algorithms that test whether graphs have a certain given property P. For example, P could be the property of being *connected*, meaning that you can get from any node to any other node by walking along a path of edges. The left graph in Figure 43 is connected whereas the right one is disconnected, since there is no way to get from nodes 1, 2 or 3 to nodes 4 or 5.

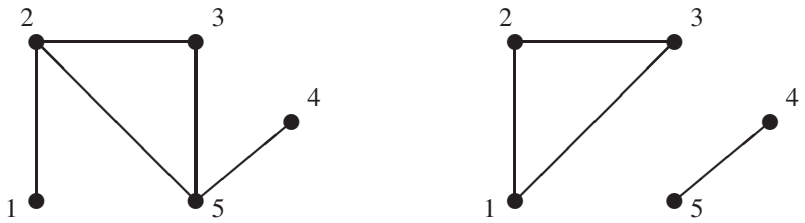


FIGURE 43
A connected and a disconnected graph

Connectedness is a very basic property of graphs which can be decided at a glance on small examples represented as a drawing. But say you have a graph with 1 million nodes, coming perhaps from a communications network or a

chip design, which is presented only as a list of edges (adjacent pairs of nodes) — then it is not quite so clear what to do if one wants to decide whether the graph is connected, making efficient use of computational resources. Among the interesting questions one can ask is whether it is possible to decide connectedness of the graph without checking for all possible pairs of nodes (there are nearly 500 billion of them) whether they are edges of the graph or not? If this were so it could conceivably lead to valuable saving of time and resources.

A basic general question to ask then is the following. Fix a positive integer n. For a given property P of graphs on n vertices, is there some algorithm that decides for every graph G on n vertices whether it has property P without knowing for every pair of nodes whether they span an edge of G or not? Think of it this way: a computer proceeds by checking whether a certain edge is in G, then it checks another edge, etc. The choice of which edge to check may depend on the outcome of the previous queries according to some definite rule (the algorithm). Is there some n-vertex graph G for which all edges must be checked before the computer can deduce whether the graph has property P or not?

If every P-testing algorithm must for at least some graph have complete knowledge about all its edges, then P is said to be an *evasive* property for n-vertex graphs. We say that a property P is *evasive* (without specifying n) if it is evasive for all $n \geq 1$.

For instance, connectedness is an evasive property. An argument showing this is given in the Appendix. A simpler property that is clearly evasive is that of having at least one edge. If a graph has no edges, then a computer cannot be sure of this until it has checked all edges, whatever algorithm it uses to generate the sequence of checked edges.

For completeness we mention without further details that nontrivial graph properties that are *not* evasive are known, a reference is given among the Notes.

The evasiveness conjecture

It has been decided for many graph properties whether they are evasive. It turns out that among the evasive ones many are *monotone*, meaning that if the property holds for some graph then it will also hold if more edges are added. For instance, connectedness is an example of a monotone property. Mounting evidence from work in the late 1960's by several researchers led to the following conjecture.

Evasiveness Conjecture. *Every monotone nontrivial graph property is evasive.*

By "nontrivial" is here meant that there is at least one graph that has the property and one that doesn't. Since monotonicity is usually easy to verify whereas evasiveness is not, this conjecture — if true — would simplify deciding evasiveness for many graph properties. Tedious case-by-case arguments, such as the one carried out for the property of connectedness in the Appendix, would not be needed.

The best general result known to date on this topic is the following theorem of Jeffry Ned Kahn, Michael Ezra Saks, and Dean Grant Sturtevant from 1984:

Kahn-Saks-Sturtevant Theorem. *The evasiveness conjecture is true for graphs on p^k nodes, for any prime number p and integer $k \geq 1$.*

This verifies the conjecture for infinitely many values of n, the number of nodes, but leaves it open when n is the product of at least two distinct primes. Thus, the smallest values of n left open are $6, 10, 12, 14, 15, \ldots$; however the case of $n = 6$ was also verified by Kahn *et al*. The general conjecture remains open, beginning with the case $n = 10$.

The proof of Kahn *et al.* makes surprising use of topology. The key idea is to view a monotone graph property for graphs on n vertices as a simplicial complex with a high degree of symmetry, to whose associated space a topological fixed point theorem can be applied. Here is how.

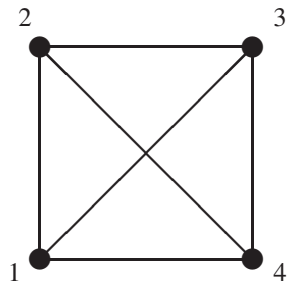


FIGURE 44
The 6 edges spanned by 4 nodes

We will keep in mind some particular monotone graph property P and consider graphs on the nodes $1, 2, \ldots, n$. Such a graph is specified by the pairs

(i,j) of nodes that are connected by an edge. Let us take the set of these pairs as the ground set for a set family Δ_n^P, whose members are the edge-sets of graphs *not* having property P. The set family Δ_n^P is closed under taking subsets, since monotonicity implies that removal of edges from a graph that doesn't have property P cannot produce a graph having that property.

Let us illustrate the idea for the case $n = 4$, taking as our monotone property connectedness. There are 6 possible edges in a graph on the nodes $1, 2, 3, 4$; see Figure 44. The simplicial complex Δ_4^{conn} of disconnected graphs on four vertices is shown in Figure 45.

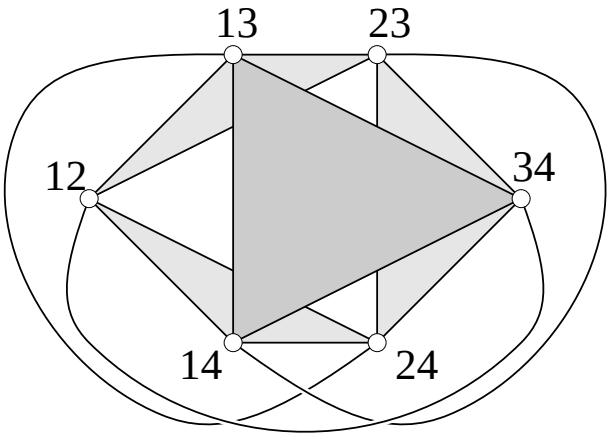
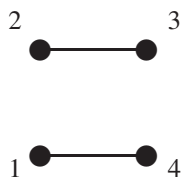
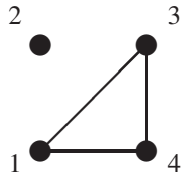


FIGURE 45
The complex of disconnected graphs on 4 nodes

In the rubber-sheet model depicted it consists of 4 triangles and 3 edges (curved line segments) glued together. To understand this picture the reader should think how to translate the vertices, edges and triangles of Δ_4^{conn} into disconnected graphs. For instance, the edge between 14 and 23 in Figure 45 corresponds to the disconnected graph



and the triangle with vertices 13, 14 and 34 corresponds to the disconnected graph



Observe in Figure 45 that the space represented by the complex Δ_4^{conn} has many "holes" — in the terminology used before this means that Δ_4^{conn} has some nonzero Betti numbers. It turns out to be a general fact, not hard to prove, that if the property P is *not* evasive then Δ_n^P is *acyclic*, meaning that all Betti numbers of Δ_n^P are equal to zero.

There are several theorems in topology to the effect that certain mappings f of an acyclic space to itself must have *fixed points*, i.e. points x such that $f(x) = x$. The best known one — one of the classics of topology — is Luitzen Egbertus Jan Brouwer's theorem from 1904, which says that every continuous mapping of an n-dimensional ball to itself has a fixed point. The one needed for the present application is a fixed point theorem of Robert Oliver from 1975, which (stripped of some technical details) says that for certain groups G of symmetry mappings of an acyclic simplicial complex Δ to itself there is a point x in the associated space such that $f(x) = x$ for *all* mappings f in G.

The complex Δ_n^P of a monotone graph property has a natural group of symmetries, namely the symmetric group S_n of all permutations of the set of nodes $1, 2, ..., n$. Permuting the nodes amounts to a relabeling (node i gets relabeled $f(i)$, etc.), and it is clear that such a relabeling will not affect whether the graph in question has property P. Therefore every permutation of $1, 2, ..., n$ induces a self-symmetry of the complex Δ_n^P of graphs not having property P.

The pieces needed for the proof of Kahn *et al.* are now at hand. Here is how they argued.

Suppose P is a monotone property for graphs on n nodes that is *not* evasive. Then, as was already mentioned, the associated complex Δ_n^P is acyclic. If furthermore $n = p^k$ then due to some special properties of prime-power numbers (the existence of finite fields) one can construct a subgroup G of S_n having the special properties needed for Oliver's fixed point theorem. Hence there is a point x in the space associated to Δ_n^P such that $f(x) = x$ for all

permutations f in G. However, this means that there is a nonempty set A in the complex Δ_n^P (that is, a graph with edge-set A not having property P) such that $f(A) = A$ for all f in G. Since G is *transitive* (meaning that if u and v are two vertices of Δ_n^P then $u = f(v)$ for some mapping f in G), A must consist of *all* vertices of Δ_n^P; that is, A is the complete graph. We have obtained that the complete graph on nodes $1, 2, \ldots, n$ does not have property P, and since P is monotone that means that *no* graph on $1, 2, \ldots, n$ can have property P, so P is trivial.

The argument shows that for monotone graph properties P on a prime-power number of nodes *nonevasive* implies *trivial*, or which is logically the same: *nontrivial* implies *evasive*.

Viewing a graph property (such as connectedness) as a simplicial complex and submitting it to topological study may seem strange. One can wonder if this point of view is of any value other than — by remarkable coincidence — for the evasiveness conjecture. It has recently become clear that this is indeed the case. Namely, the complexes Δ_n^{conn} of disconnected graphs on n vertices have arisen and play a role in the work of Victor Anatol'evich Vassiliev on knot invariants. Also some other monotone graph properties have naturally presented themselves as simplicial complexes in other mathematical contexts.

APPENDIX

To see that connectedness is an evasive property we can argue as follows. Imagine that we have a computer running a program that tests graphs for connectedness. The graphs to be tested, whose nodes we may assume are labeled $1, 2, ..., n$, are presented to the computer in the form of an $n \times n$ upper-triangular matrix of zeros and ones, with a 1 entry in row i and column j, for $i < j$, if (i, j) is an edge of the graph and a 0 entry otherwise. For instance, here are the matrices representing the graphs in Figure 43:

$$\begin{array}{ccccc} * & 1 & 0 & 0 & 0 \\ & * & 1 & 0 & 1 \\ & & * & 0 & 1 \\ & & & * & 1 \\ & & & & * \end{array} \qquad \begin{array}{ccccc} * & 1 & 1 & 0 & 0 \\ & * & 1 & 0 & 0 \\ & & * & 0 & 0 \\ & & & * & 1 \\ & & & & * \end{array}$$

The computer is allowed to inspect only one entry of this matrix at a time, and what we want to show is that for some graph it must in fact inspect all of them. To find such a worst-case graph we can imagine playing the

following game with the computer. Say that instead of deciding on the graph in advance, we write the zeros and ones (specifying its nonedges and edges) into the matrix only at the last moment, as the computer demands to inspect them. Say furthermore that we do this according to the following strategy (designed to keep the computer making as many queries as possible): When the computer goes to inspect the (i,j) entry of the matrix (according to whatever algorithm it is using), then

- write 0 into position (i,j) if it is not possible to conclude from the partial information known to the computer at that time — including this last 0 — that the graph is disconnected,
- otherwise, write 1 into position (i,j).

It is an elementary but somewhat tricky argument to show that this strategy will force the computer to inspect all entries of the matrix before it can decide whether the corresponding graph is connected or not. We will outline a proof by finite induction.

The crucial step is to prove the following claim:

Suppose that at some stage 1 *is written into position* (i,j). *Let* A *be the set of nodes that are at that stage connected to* i *by* 1*-marked edges (including* i *itself), and let* B *be the set of nodes connected to* j *by* 1*-marked edges (including* j*). Then after* 1 *has been written into position* (i,j)*, all possible edges between nodes in* $A \cup B$ *have been inspected.*

(Clarification: "at that stage" refers to the configuration existing at the time just before 1 is assigned to the position/edge (i,j). At that time some other edges have already been inspected and are marked with 0 or 1, while the remaining have not yet been inspected.)

Note that $A \cap B = \emptyset$, and that $|A \cup B| \geq 2$ since $i \in A$ and $j \in B$. The statement is clearly true if $|A \cup B| = 2$, and we proceed by induction on the number of elements of $A \cup B$.

Suppose that $|A \cup B| > 2$. Since what is written into position (i,j) is *not* 0, that means that there is some partition $C \cup D = \{1,2,\ldots,n\}$ into nonempty disjoint subsets C and D such that $i \in C$, $j \in D$ and all possible edges $\{c,d\} \neq \{i,j\}$ with $c \in C$ and $d \in D$ are already marked with 0. Clearly, we must have $A \subseteq C$ and $B \subseteq D$, so in particular all edges between a node in A and a node in B have already been inspected. Also, all edges between two nodes both in A have by the induction assumption been inspected, and similarly for B. This covers all possible edges between nodes in $A \cup B$ and the claim follows.

Suppose now that connectedness/disconnectedness can be decided after inspection of k matrix entries, and that k is the minimum such number. According to our strategy for writing 0 or 1, the outcome can never be that the graph is disconnected. Also, if the kth entry is 0 and the graph is connected we have a contradiction, since then the information needed to conclude connectedness would have been available already before the kth entry was inspected. So, the kth entry is 1, and since the conclusion is that the graph is connected the claim above implies that all other entries have already been inspected before the kth one. This proves that connectedness is an evasive graph property.

17
Notes

Here we give some hints for further reading. We refer mainly to general accounts that should be at least partially accessible to the non-expert and that give lots of further references. Original sources are only rarely mentioned.

For a broad view of current combinatorics, with a wealth of information and references (but partly written at a high level), see the *Handbook of Combinatorics* [1].

CHAPTERS 1–4

The basic theory of enumeration is developed in the books [38, 39]. The combinatorics of number and set partitions, standard Young tableaux, and generating functions, together with algebraic ramifications, is discussed there. A briefer account of this material is given in [25]. More elementary references on enumeration are [15, 16]. A nice introduction to generating functions is given in [44]. The web site [37] is a valuable companion to the study of enumeration.

CHAPTER 2

Two good references for number partitions are the books [3, 4], of which the second is more elementary.

CHAPTERS 3–5

An introduction to the combinatorics of permutations and Young tableaux can be found in [39, Ch. 7], as well as in [31, Ch. 5.1] and in [36]. The latter book also gives an accessible introduction to the connections with representation theory. For the connections with representation theory and algebraic geometry, see [23].

CHAPTER 6

A survey of the theory of increasing and decreasing subsequences is given in [41]. More about the application to airplane boarding can be found in [8], [9], [10].

CHAPTER 7

Reduced decompositions of permutations are part of a more general scenario provided by the theory of Coxeter groups. See Chapters 3 and 7 of [13]. The results we have discussed first appeared in [40]. Another approach to this subject is given in [24].

CHAPTERS 8–9

There is a huge literature on tilings, but most of this is not concerned with enumerative problems. For a wealth of information concerning the non-enumerative aspects see the book [27]. At present there is no good introduction to the enumerative aspects of tilings. The results that we have mentioned can be found in the references [18, 19, 20, 28, 29]. A general survey of planar tilings for a nonmathematical audience is given by [5].

CHAPTER 10

For more about the Möbius function see [38, Ch. 3].

CHAPTER 11

The combinatorial theory of hyperplane arrangements is exposited in [42], and is placed in a broader mathematical context in [33]. See also [11].

Several bijections have been given between the regions of the Shi arrangement and trees (or combinatorial objects with well-known bijections with trees). A nice such bijection with additional references is given in [7].

CHAPTER 12

For a general introduction to convex polytopes and an accessible discussion of the g-theorem we refer to [45]. A classic in the field is [26]. Figure 31 is due to K. Fukuda.

CHAPTER 13

Connections between combinatorics and topology are discussed in the book [34] and in the survey [12]. There are several books giving a gentle introduction to topology, see e.g. [6].

CHAPTER 14

Information about extremal set-theoretic combinatorics can be gotten from [14], [21] and parts of [2] and [43]. For more about Frankl's conjecture see [17], for extensions of the Erdős-Ko-Rado Theorem see [22] and for Kneser's conjecture see [34]. Borsuk's problem is discussed in [2] and [35].

CHAPTERS 15–16

Connections between combinatorics and computer science is a huge subject that we have barely touched. For some glimpses see [32], and for sorting algorithms also [31]. More about the k-equal problem and its solution can be found in [11], while more details about the evasiveness conjecture can be found in [12, 14]. For an example of a nonevasive nontrivial graph property, see [14, p. 1284]. Various monotone graph properties seen from a topological point of view are discussed in [30].

Bibliography

[1] *Handbook of Combinatorics* (R. Graham, M. Grötschel and L. Lovász, eds.), North-Holland, Amsterdam/New York, and MIT Press, Cambridge, Massachusetts, 1995.

[2] M. Aigner and G. M. Ziegler, *Proofs from The Book, Third ed.*, Springer-Verlag, Berlin/Heidelberg, 2004.

[3] G. E. Andrews, *The Theory of Partitions*, Encyclopedia of Mathematics and Its Applications, Vol. 2, Addison-Wesley, Reading, Massachusetts, 1976.

[4] G. E. Andrews and K. Eriksson, *Integer Partitions*, second revised ed., Cambridge University Press, Cambridge/New York, 2004.

[5] F. Ardila and R. Stanley, Tilings, to appear in *Math. Intelligencer*. Available electronically at
`www.claymath.org/fas/senior_scholars/Stanley/tilings.pdf`

[6] M. A. Armstrong, *Basic Topology*, Springer, 1983.

[7] C. A. Athanasiadis and S. Linusson, A simple bijection for the regions of the Shi arrangement of hyperplanes, *Discrete Math.* **204** (1999), 27–39.

[8] E. Bachmat, E., D. Berend, L. Sapir, S. Skiena, and N. Stolyarov, Analysis of airline boarding via space-time geometry and random matrix theory, *J. Physics A : Mathematical and General* **39** (2006), L453–L459.

[9] E. Bachmat, E., D. Berend, L. Sapir, and S. Skiena, Optimal boarding policies for thin passengers, *Adv. in Appl. Probab.* **39** (2007), 1098–1114.

[10] E. Bachmat, E., D. Berend, L. Sapir, S. Skiena, and N. Stolyarov, Analysis of airplane boarding times, *Operations Research* **57** (2009), 499–513.

[11] A. Björner, Subspace arrangements, in *First European Congress of Mathematics, Paris 1992* (A. Joseph *et al.*, eds.), Progress in Mathematics Series, Vol. 119, Birkhäuser, Boston, 1994, pp. 321–370.

[12] A. Björner, Topological methods, pp. 1819–1872 in *Handbook of Combinatorics*.

[13] A. Björner and F. Brenti, *Combinatorics of Coxeter Groups*, GTM series Vol. 231, Springer-Verlag, NewYork/Berlin/Heidelberg, 2005.

[14] B. Bollobás, Extremal graph theory, pp. 1231–1292 in *Handbook of Combinatorics*.

[15] M. Bóna, *Combinatorics of Permutations*, Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[16] M. Bóna, *Introduction to Enumerative Combinatorics*, McGraw-Hill, New York, 2005.

[17] I. Bosnjak and P. Marković, The 11-element case of Frankl's conjecture, *Electronic J. Combinatorics* **15** (2008), # R88.

[18] H. Cohn, N. Elkies, and J. Propp, Local statistics for random domino tilings of the Aztec diamond, *Duke Math. J.* **85** (1996), 117–166.

[19] M. Ciucu, Enumeration of perfect matchings in graphs with reflective symmetry, *Journal of Combinatorial Theory, Series A* **77** (1997), 67–97.

[20] N. Elkies, G. Kuperberg, M. Larsen, and J. Propp, Alternating-sign matrices and domino tilings, Parts I and II, *Journal of Algebraic Combinatorics* **1** (1992), 111–132, 219–234.

[21] P. Frankl, Extremal set systems, pp. 1293–1329 in *Handbook of Combinatorics*.

[22] P. Frankl and Z. Füredi, Extremal problems concering Kneser graphs, *Journal of Combinatorial Theory, Series B* **40** (1986), 270–284.

[23] W. Fulton, *Young Tableaux: With Applications to Representation Theory and Geometry*, Cambridge University Press, Cambridge, 1997.

[24] A. M. Garsia, The Saga of Reduced Decompositions of Elements of the Symmetric Group, Laboratoire de combinatoire et d'informatique mathématique (LACIM), no. 29, Université du Québec à Montréal, Montréal, 2002.

[25] I. Gessel and R. P. Stanley, Algebraic Enumeration, pp. 1021–1061 in *Handbook of Combinatorics*.

[26] B. Grünbaum, *Convex Polytopes, Second ed.*, GTM series Vol. 221, Springer-Verlag, Berlin, 2003.

[27] B. Grünbaum and G. C. Shephard, *Tilings and Patterns*, Freeman, New York, 1987.

[28] W. Jockusch, Perfect matchings and perfect squares, *Journal of Combinatorial Theory, Series A* **67** (1994), 100–115.

[29] W. Jockusch, J. Propp, and P. Shor, Random domino tilings and the Arctic circle theorem, preprint; `math.CO/9801068`.

[30] J. Jonsson, *Simplicial complexes of graphs*, Lecture Notes in Math., Vol. 1928, Springer, 2008.

[31] D. E. Knuth, *The Art of Computer Programming, Vol. 3*, Addison-Wesley, Reading, Massachusetts, 1973 (updated and reprinted 1997).

[32] L. Lovász, D. B. Shmoys, and É. Tardos, Combinatorics in Computer Science, pp. 2003–2038 in *Handbook of Combinatorics*,

[33] P. Orlik and H. Terao, *Arrangements of Hyperplanes*, Springer-Verlag, Berlin/Heidelberg, 1992.

[34] J. Matousek, *Using the Borsuk-Ulam theorem*, Springer-Verlag, Berlin/Heidelberg, 2003.

[35] A. M. Raigorodskii, The Borsuk partition problem: The seventieth anniversary, *Math. Intelligencer* **26** (2004), 4–12.

[36] B. E. Sagan, *The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions*, second ed., Springer-Verlag, New York/Berlin/Heidelberg, 2001.

[37] N. J. A. Sloane and S. Plouffe, *The Encyclopedia of Integer Sequences*, `www.research.att.com/~njas/sequences`.

[38] R. P. Stanley, *Enumerative Combinatorics*, Volume 1, Wadsworth & Brooks/Cole, Monterey, CA, 1986 (second printing, Cambridge University Press, Cambridge/New York, 1997).

[39] R. P. Stanley, *Enumerative Combinatorics*, Volume 2, Cambridge University Press, Cambridge/New York, 1999.

[40] R. P. Stanley, On the number of reduced decompositions of elements of Coxeter groups, *European J. Combinatorics* **5** (1984), 359–372.

[41] R. P. Stanley, Increasing and decreasing subsequences and their variants, *Proc. Internat. Cong. Math. (Madrid, 2006)*, vol. 1, American Mathematical Society, Providence, RI, 2007, pp. 545–579.

[42] R. P. Stanley, An introduction to hyperplane arrangements, in *Geometric Combinatorics* (E. Miller, V. Reiner, and B. Sturmfels, eds.), IAS/Park City Mathematics Series Vol.13, American Mathematical Society, Providence, RI, 2007, pp. 389–496.

[43] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics, Second ed.*, Cambridge Univ. Press, Cambridge, UK, 2001.

[44] H. S. Wilf, *generatingfunctionology*, Academic Press, San Diego, 1990.

[45] G. M. Ziegler, *Lectures on Polytopes*, GTM Series Vol. 152, Springer-Verlag, NewYork/Berlin/Heidelberg, 1995.

Cited mathematicians

Harold Neville Vazeille Temperley, b. 1915 67
Robert McDowell Thrall, 1914–2006 36
Victor Anatol'evich Vassiliev, b. 1956 148
Anatoly Moiseevich Vershik, b. 1933 53
Julian West, b. 1964 62
Richard Michael Wilson, b. 1945 131
Andrew Chi-Chih Yao, b. 1946 139
Alfred Young, 1873–1940 38
Thomas Zaslavsky, b. 1945 96