A Combinatorial Miscellany

Anders Björner[1]
Matematiska Institutionen
Kungl. Tekniska Högskolan
S-100 44 Stockholm
SWEDEN

Richard P. Stanley[2]
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
U.S.A.

September 13, 2004

1 Introduction.

A recent newcomer to the 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

[1]Partially supported by the Mathematical Sciences Research Institute (Berkeley, CA).
[2]Partially supported by NSF grant #DMS-9500714.

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 article 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 both 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. 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 suggestions for further reading point 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, acb, bac, bca, cab, 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. 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 also illustrate (in Section 11) 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.

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.

Our final topic is a glimpse of 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.

To understand this article should for the most part require hardly any knowledge of mathematics beyond high-school algebra. Only some details in the boxes and in the last few sections (having to do with topology) are a bit more demanding.

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. 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. A partition is allowed to have only one part (summand), so that 5 is a partition of 5. There are in fact seven partitions of 5, given by

$$\begin{array}{l} 5 \\ 4+1 \\ 3+2 \\ 3+1+1 \\ 2+2+1 \\ 2+1+1+1 \\ 1+1+1+1+1. \end{array}$$

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 (1905–1991) computed $p(14,031)$ (a number with 127 digits!), and nowadays a computer would have no trouble computing $p(10^{12})$, a number with 1,113,996 digits. 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

$$F(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots,$$

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 $G(x) = b_0 + b_1 x + b_2 x^2 + \cdots$, 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 $a_i x^i$ from $F(x)$ and a term $b_j x^j$ from $G(x)$ and multiply them to get $a_i b_j x^{i+j}$. We then add together all such terms. For instance, the term in the product involving x^4 will be

$$\begin{aligned} a_0 \cdot b_4 x^4 + a_1 x \cdot b_3 x^3 + a_2 x^2 \cdot b_2 x^2 + a_3 x^3 \cdot b_1 x + a_4 x^4 \cdot b_0 \\ = (a_0 b_4 + a_1 b_3 + a_2 b_2 + a_3 b_1 + a_4 b_0) x^4. \end{aligned}$$

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

$$a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_{n-1} b_1 + a_n b_0.$$

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 $a_0 b_0 = 1 \cdot 1 = 1$. If $n > 1$ then the coefficient of x^n is $a_n b_0 + a_{n-1} b_1 = 1 - 1 = 0$ (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

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

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

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

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

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

Although equation (3) is very elegant, one may ask whether it is of any use. Can it be used to obtain interesting information about the numbers $p(n)$? We will first 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 (3) was obtained we get

$$R(x) = \frac{1}{(1-x)(1-x^3)(1-x^5)(1-x^7)\cdots}. \tag{4}$$

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 (3) will have no trouble seeing that

$$Q(x) = (1+x)(1+x^2)(1+x^3)\cdots. \tag{5}$$

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 (5) we obtain

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

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 (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, 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 *one-to-one correspondence* or *bijection* between the partitions of n into odd parts and the partitions of n into distinct parts. Such a bijection would yield a *bijective proof* of the formula $q(n) = r(n)$. Exhibiting a bijection between two different (finite) sets is considered the most elegant and natural way to show that they have the same number of elements. Such bijective proofs can involve considerable ingenuity, while the method of generating functions often yields a more mechanical proof technique.

We now would like to give a bijective proof of Euler's formula $q(n) = r(n)$. Several such proofs are known; we give the perhaps simplest of these, due to James Joseph Sylvester (1814–1897). 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

$$\begin{aligned} 202 = 19 + 19 + 19 + 11 + 11 + 11 + 11 + 9 + 7 + 7 + 7 + 5 \\ +5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 1 + 1 + 1 + 1 + 1 + 1. \end{aligned}$$

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):

$$202 = (38 + 19) + 44 + 9 + (14 + 7) + (40 + 20 + 5) + (4 + 2). \qquad (7)$$

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 (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.

The subject of partitions is replete with results similar to Euler's, in which two sets of partitions have the same number of elements. The most famous of these results is called the *Rogers-Ramanujan identities*, after Leonard James Rogers (1862–1933) and Srinivasa Aiyangar Ramanujan (1887–1920), who proved these identities in the form of an identity between generating functions. It was Percy Alexander MacMahon (1854–1929) 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, $g(13) = 10$:

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

$$= 6+1+1+1+1+1+1+1 = 4+4+4+1 = 4+4+1+1+1+1+1$$

$$= 4+1+1+1+1+1+1+1+1+1 = 1+1+1+1+1+1+1+1+1+1+1+1+1.$$

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. For a long time it was an open problem to find a bijective proof of the Rogers-Ramanujan identities, but such a proof was finally given in 1980 by Adriano M. Garsia (b. 1928) and Stephen Carl Milne (b. 1949). However, their proof is very complicated, and it would still be of great interest to find a simple, conceptual bijective proof.

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 (b. 1940) using the Rogers-Ramanujan identities.

$$
\begin{array}{llllllllll}
7&4&4&4&2&2&1&1&1&1\\
7&4&4&2&2&1&1&1&1\\
6&3&2&2&2&1&1&1&1\\
4&2&2&1&1&1\\
2&2&1&1&1\\
2&1&1&1&1\\
1&1&1&1&1\\
1&1
\end{array}
$$

Figure 1: A plane partition

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. 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 1.

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 1 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 1, 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

$$\begin{array}{ccccccccccc} \emptyset & 1 & 2 & 11 & 1 & 3 & 21 & 111 & 11 & 2 & 1 \\ & & & & 1 & & & & 1 & 1 & 1 \\ & & & & & & & & & & 1 \end{array} .$$

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. Then MacMahon established the remarkable formula

$$\begin{aligned} & pp(0) + pp(1)x + pp(2)x^2 + pp(3)x^3 + \cdots \\ & \quad = \frac{1}{(1-x)(1-x^2)^2(1-x^3)^3(1-x^4)^4\cdots}. \end{aligned} \tag{8}$$

Unlike Euler's formula (3) for the generating function for the number $p(n)$ of ordinary partitions of n, MacMahon's 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 (b. 1942) and Donald Ervin Knuth (b. 1938). Their proof was based on the *Schensted correspondence*, 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 (1906–19??) in a rather vague form in 1938 (with some assistance from Dudley Ernest Littlewood (1903–1979)), and later more explicitly by Craige Eugene Schensted (b. 19??) in 1961. Schensted's motivation for looking at this correspondence is discussed in Section 5. The version of Schensted's correspondence used here is due to Knuth.

We now give a brief account of the proof of Bender and Knuth. Using equation (1), the product on the right-hand side of (8) may be written

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

$$(1+x^2+x^4+\cdots)(1+x^3+x^6+\cdots)(1+x^3+x^6+\cdots)(1+x^3+x^6+\cdots)\cdots. \quad (9)$$

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 (8) 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 arry.* A typical two-line array A looks like

$$A = \begin{matrix} 3\ 3\ 3\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1 \\ 3\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 4\ 4\ 3\ 3\ 3 \end{matrix}. \quad (10)$$

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 (9) as follows. Let a_{ij} be the number of columns $\substack{i\\j}$ of A. For instance (always referring to the two-line array (10)), $a_{33}=1,\ a_{31}=2,\ a_{13}=3,\ a_{23}=0$. Given a_{ij}, let $k=i+j-1$. Then choose the term $x^{a_{ij}\cdot k}$ from the ith factor of (9) 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 (9) (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 Schensted correspondence. It will be described as an algorithm that we call the *Schensted 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 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 just above a in A) 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 2 illustrates the bumping procedure with the two-line array A of equation (10). 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 2 we obtain:

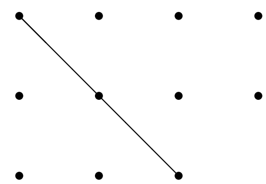
$$P = \begin{array}{llllll} 4&4&3&3&3&1 \\ 3&2&2&2&1 & \\ 1&1 & & & & \end{array}, \qquad Q = \begin{array}{llllll} 3&3&3&2&2&2 \\ 2&2&1&1&1 & \\ 1&1 & & & & \end{array}. \tag{11}$$

The final step of the Bender-Knuth bijection is to merge the two column-

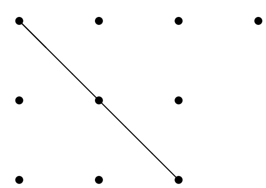
i	P_i	Q_i
1	3	3
2	3 1	3 3
3	3 1 1	3 3 3
4	3 2 1 1	3 3 3 2
5	3 2 2 1 1	3 3 3 2 2
6	3 2 2 2 1 1	3 3 3 2 2 2
7	3 2 2 2 1 1 1	3 3 3 2 2 2 2
8	3 2 2 2 1 1 1 1	3 3 3 2 2 2 2 2
9	4 2 2 2 1 1 3 1 1	3 3 3 2 2 2 2 2 1
10	4 4 2 2 1 1 3 2 1 1	3 3 3 2 2 2 2 2 1 1
11	4 4 3 2 1 1 3 2 2 1 1	3 3 3 2 2 2 2 2 1 1 1
12	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 2: The Schensted correspondence

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 (11). 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

$$\pi = \begin{array}{llllll} 4 & 4 & 3 & 3 & 3 & 1 \\ 4 & 3 & 3 & 3 & 2 & 1 \\ 3 & 3 & 1 & & & \end{array}\ . \tag{12}$$

This gives the desired bijection that proves MacMahon's formula (8). 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*. A good way to convince yourself of the accuracy of the procedure is to take the plane partition π of equation (12) and try to reconstruct the original choice of terms from the product $1/(1-x)(1-x^2)^2\cdots$.

By analyzing more carefully the above bijective proof, it is possible to extend the formula (8) 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

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

$$\frac{1}{[1][2]^2[3]^3 \cdots [r]^r[r+1]^r \cdots [s]^r[s+1]^{r-1}[s+2]^{r-2} \cdots [r+s-1]}. \tag{13}$$

For instance, when $r = 3$ and $s = 5$ the right-hand side of equation (13) 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

$$\begin{array}{cccccccccccc} 4 & 31 & 22 & 211 & 1111 & 3 & 2 & 21 & 11 & 111 & 2 & 11 \\ & & & & & 1 & 2 & 1 & 11 & 1 & 1 & 1 \\ & & & & & & & & & & 1 & 1 \end{array}.$$

By more sophisticated arguments (not a direct bijective proof) one can extend equation (13) even further, as follows. Let $pp_{rst}(n)$ 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

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

$$\frac{[1+t][2+t]^2[3+t]^3 \cdots [r+t]^r[r+1+t]^r \cdots [s+t]^r[s+1+t]^{r-1}[s+2+t]^{r-2} \cdots [r+s-1+t]}{[1][2]^2[3]^3 \cdots [r]^r[r+1]^r \cdots [s]^r[s+1]^{r-1}[s+2]^{r-2} \cdots [r+s-1]}. \tag{14}$$

Note that the right-hand sides of equations (13) and (14) have the same denominator. The numerator of (14) is obtained by replacing each denominator factor $[i]$ with $[i+t]$. Equation (14) was also first proved by MacMahon, and is the culmination of his work on plane partitions. It is also closely related to the *representation theory of Lie groups and Lie algebras*, a subject that at first sight has no connection with plane partitions. (See Box.) 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 analogues.

As a concrete example of equation (14), suppose that $r = 2$, $s = 3$, and $t = 2$. The right-hand side of (14) becomes

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

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

The Schensted correspondence has a number of remarkable properties that were not needed for the derivation of MacMahon's formula (8). The most striking of these properties is the following. Consider a two-line array A such as (10) which is the input to the Schensted correspondence. 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'. For the two-line array A of equation (10) we have

$$A' = \begin{matrix} 4 & 4 & 3 & 3 & 3 & 3 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 2 & 2 \end{matrix} \tag{15}$$

Thus the Schensted correspondence can be applied to A'. If (P, Q) is the pair of column-strict plane partitions obtained by applying the Schensted correspondence to A, then applying this correspondence 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 (10) and (15) that we obtain such a simple result. We can use this "symmetry property" of the Schensted correspondence 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

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

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

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

Without going into any details, let us just say that the symmetry property of the Schensted correspondence just described yields a bijective proof, similar to the proof we have given of MacMahon's formula (8), of the generating function

$$s(0) + s(1)x + s(2)x^2 + \cdots$$
$$= \frac{1}{(1-x)(1-x^3)(1-x^4)(1-x^5)(1-x^6)(1-x^7)(1-x^8)^2(1-x^9)(1-x^{10})^2\cdots}.$$

The exponent of $1-x^{2k-1}$ in the denominator 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

$$\begin{array}{l} 1\,3\,4\,6 \\ 2\,7\,8 \\ 5\,9 \end{array} \tag{16}$$

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

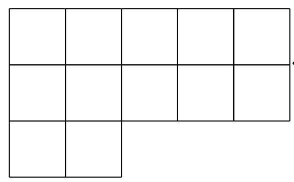
$$\begin{array}{llllllllll} 1\,2\,3\,4 & 1\,2\,3 & 1\,2\,4 & 1\,3\,4 & 1\,2 & 1\,3 & 1\,2 & 1\,3 & 1\,4 & 1 \\ & 4 & 3 & 2 & 3\,4 & 2\,4 & 3 & 2 & 2 & 2 \\ & & & & & & 4 & 4 & 3 & 3 \\ & & & & & & & & & 4 \end{array}$$

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 (16) corresponds to the ballot sequence $1, 2, 1, 1, 3, 1, 2, 2, 3$.

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

$$\begin{array}{lllllllll} 1\,2\,3\,4 & 1\,2\,3\,5 & 1\,2\,3\,6 & 1\,2\,4\,5 & 1\,2\,4\,6 & 1\,2\,5\,6 & 1\,3\,4\,5 & 1\,3\,4\,6 & 1\,3\,5\,6 \\ 5\,6 & 4\,6 & 4\,5 & 3\,6 & 3\,5 & 3\,4 & 2\,6 & 2\,5 & 2\,4 \end{array}$$

A formula for f^λ (stated in terms of ballot sequences) was given by MacMahon in 1900. A simplified version was given was James Sutherland Frame (1907–1997), Robinson (mentioned earlier in connection with the Schensted correspondence), and Robert McDowell Thrall (b. 1914) 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 hook-length formula states that

$$f^\lambda = \frac{n!}{H_\lambda}, \tag{17}$$

where λ is a partition of n and $n!$ (read "n factorial") is short for $1 \cdot 2 \cdots n$. 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, but none is as simple as the proof we have sketched of equation (8) using Schensted's correspondence.

In addition to their usefulness in combinatorics, SYT also play a significant role in the theory of symmetry. This important theory was developed primarily by Alfred Young (1873–1940), 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 the Box entitled "Connections with representation theory" for more details. An immediate consequence of this theory is that the number of ordered pairs of SYT of the same shape and with n squares is equal to $n!$, the number of permutations of n objects. For instance, when $n = 3$ we get the six pairs

$$\left(\begin{matrix}123 & 123\end{matrix}\right) \quad \left(\begin{matrix}12 & 12\\ 3 & 3\end{matrix}\right) \quad \left(\begin{matrix}12 & 13\\ 3 & 2\end{matrix}\right)$$

$$\left(\begin{matrix}13 & 12\\ 2 & 3\end{matrix}\right) \quad \left(\begin{matrix}13 & 13\\ 2 & 2\end{matrix}\right) \quad \left(\begin{matrix}1 & 1\\ 2 & 2\\ 3 & 3\end{matrix}\right).$$

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

$$\sum_{\lambda \vdash n} \left(f^\lambda\right)^2 = n!, \tag{18}$$

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 $1, 2, \ldots, n$, which may be regarded as simply a way of listing them in some order, such as $5, 2, 7, 6, 1, 4, 3$ (or just 5276143 when no confusion can arise), can we associate with w a pair (T_1, T_2) 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 Schensted correspondence! There is only one

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

Figure 3: An SYT and its corresponding reverse SYT

minor technicality that needs to be explained before we apply the Schensted correspondence. Namely, the column-strict plane partitions we were dealing 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 3.

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

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

When we apply the Schensted correspondence 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

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

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

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

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. (The details of this argument are left as an exercise.) 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 (18). This remarkable connection between permutations and tableaux is the foundation for an elaborate theory of permutation enumeration. In the next section we give a taste of this theory.

BOX: Connections with representation theory. In this box we assume familiarity with the fundamentals of representation theory. First we consider the group $G = \mathrm{GL}(n, \mathbb{C})$ 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 $\varphi : G \to \mathrm{GL}(N, \mathbb{C})$, 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 $\varphi : \mathrm{GL}(2, \mathbb{C}) \to \mathrm{GL}(3, \mathbb{C})$ defined by

$$\varphi \begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \begin{bmatrix} a^2 & 2ab & b^2 \\ ac & ad + bc & bd \\ c^2 & 2cd & d^2 \end{bmatrix} \tag{19}$$

preserves multiplication (and the identity element), and hence is a polynomial representation of $\mathrm{GL}(2, \mathbb{C})$ of degree 3. Let $\varphi : \mathrm{GL}(n, \mathbb{C}) \to \mathrm{GL}(N, \mathbb{C})$ 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 (19) 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 (19), we have

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

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

- Every polynomial representation (assumed finite-dimensional) of the group $\mathrm{GL}(n, \mathbb{C})$ is completely reducible, i.e., a direct sum of irreducible polynomial representations. These irreducible constituents are unique up to equivalence.

- 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.

The effect of these properties is that once we determine the character of a polynomial representation φ of $\mathrm{GL}(n, \mathbb{C})$, 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 character of $\mathrm{GL}(n, \mathbb{C})$. The main result is the following.

Fundamental theorem on the polynomial characters of GL$(n, \mathbb{C})$. *The irreducible characters of* $\mathrm{GL}(n, \mathbb{C})$ *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 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),

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

This is just the character of the representation defined by equation (19). Hence this representation is one of the irreducible representations of $\mathrm{GL}(2,\mathbb{C})$.

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

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

Hence

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

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 $\mathrm{GL}(n,\mathbb{C})$, when restricted to $\mathrm{SL}(n,\mathbb{C})$, is just the adjoint representation of $\mathrm{SL}(n,\mathbb{C})$.)

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 $\mathrm{GL}(n,\mathbb{C})$ and $\mathfrak{S}_k$ discovered by Issai Schur (1875–1941).

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

$$V^{\otimes k} = \coprod_\lambda \left(M^\lambda \otimes F_\lambda \right), \tag{20}$$

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 $\mathrm{GL}(n,\mathbb{C})$-module corresponding to λ, and M^λ is an irreducible $\mathfrak{S}_k$-module. Thus when

$k \leq n$, 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 Schensted correspondence (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. Comparing with equation (20) and using the fact that the character of $\mathrm{GL}(n, \mathbb{C})$ acting on $V^{\otimes k}$ is $(x_1 + \cdots + x_n)^k$, we see 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 (18) (with n replaced by k).

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

5 Increasing and decreasing subsequences.

In this section we discuss an unexpected connection between the Schensted correspondence 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.

We will be interestested in the length of the *longest* increasing and decreasing subsequences of a permutation w. Denote by $i(w)$ the length of the longest increasing subsequence of w, and by $d(w)$ the length of the longest decreasing subsequence. By careful inspection one sees for instance that $i(6251743) = 3$ and $d(6251743) = 4$. It is intuitively plausible that there should be some kind of tradeoff between the values $i(w)$ and $d(w)$. If $i(w)$ 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 $d(w)$ to be large. An extreme case occurs when $i(w) = 1$. Then there is only one choice for w, namely, $n, n - 1, \ldots, 1$, and we have $d(w) = n$.

How can we quantify the feeling that that $i(w)$ and $d(w)$ cannot both be small? A famous result of Pal (??) Erdős (1913–1996) and George Szekeres (b. 1911), 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*. Let w be a permutation of $1, 2, \ldots, n$. The Erdős-Szekeres theorem states that if p and q are positive integers for which $n > pq$, then either $i(w) > p$ or $d(w) > q$. Moreover, this result is best possible in the sense that if $n = pq$ then we can find at least one permutation w such that $i(w) = p$ and $d(w) = q$. An equivalent way to formulate the Erdős-Szekeres theorem is by the inequality

$$i(w) \cdot d(w) \geq n,$$

showing clearly that $i(w)$ and $d(w)$ cannot both be small. For instance, both can't 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 (1916–1988) 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 $w = a_1a_2 \cdots a_n$ 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 $i(w) > p$. Similarly, if $s_i > q$ then $d(w) > q$, 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 $i(w) = p$ and $d(w) = q$. It is easy to check that the following choice of w works:

$$w = (q-1)p+1, (q-1)p+2, \ldots, qp, (q-2)p+1, (q-2)p+2, \ldots, (q-1)p, \\ \ldots, 2p+1, 2p+2, \ldots, 3p, p+1, p+2, \ldots, 2p, 1, 2, \ldots, p. \qquad (21)$$

This completes the proof of the Erdős-Szekeres theorem.

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 $i(w) = p$ and $d(w) = q$, 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 Schensted correspondence on the other.

There are two fundamental properties of the Schensted correspondence that are needed for our purposes. Suppose we apply the Schensted correspondence to a permutation $w = a_1 a_2 \cdots a_n$ of $1, 2, \ldots, n$, getting two column-strict plane partitions P and Q whose parts are $1, 2, \ldots, n$. The first property we need of the Schensted correspondence 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

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

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 Box for a proof by induction of Property 1.

BOX. *Proof of Property 1.* Recall that $w = a_1a_2\cdots a_n$. We prove by induction on j that after the Schensted algorithm has been applied to $a_1a_2\cdots a_j$, 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 $a_1a_2\cdots a_j$ 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 true for j. Suppose that the first row of P_j is $c_1c_2\cdots c_r$. By the induction hypothesis, c_i is the rightmost term of the sequence $a_1a_2\cdots a_j$ 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 Schensted 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 $a_1a_2\cdots a_{j+1}$ ending at a_{j+1} has length i, since clearly a_{j+1} will be the *rightmost* element of $a_1a_2\cdots a_{j+1}$ 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 $a_1a_2\cdots a_{j+1}$, 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 $a_1a_2\cdots a_j$ 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 Schensted 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.

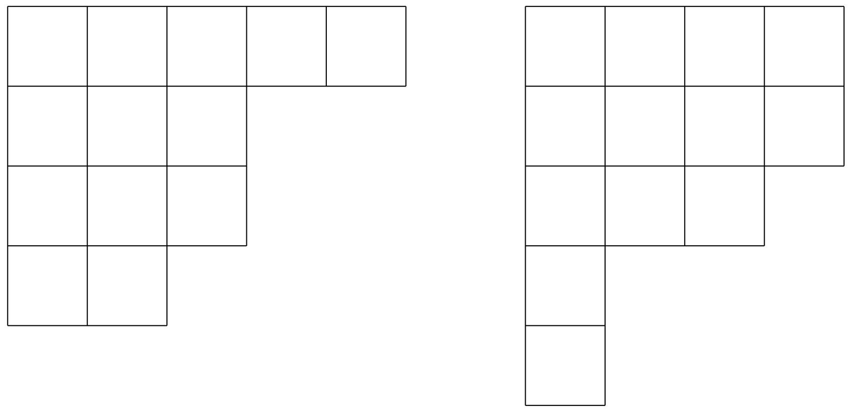


Figure 4: The Young diagram of a partition and its conjugate

The second property we need of the Schensted correspondence 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 4.

Property 2. Suppose that when the Schensted correspondence is applied to a permutation $w = a_1 a_2 \cdots a_n$, we obtain the pair (P, Q) of reverse SYT. Let $\bar{w} = a_n a_{n-1} \cdots a_1$, the *reverse* permutation of w. Suppose that when the Schensted correspondence is applied to $\bar{w}$, we obtain the pair $(\bar{P}, \bar{Q})$ of reverse SYT. Then the shape of $\bar{P}$ (or $\bar{Q}$) 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 $\bar{P}$ is actually the *transpose* of P, obtained by interchanging the rows and columns of P. (The connection between Q and $\bar{Q}$ 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 Schensted correspondence to the permutation w and get a pair (P,Q) of reverse SYT of shape $\lambda = (\lambda_1, \lambda_2, \ldots)$, then Property 1 tells us that

$$d(w) = \lambda_1.$$

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 Schensted correspondence to the reverse permutation $\bar{w}$, obtaining the pair $(\bar{P}, \bar{Q})$ of reverse SYT. When we reverse a permutation, increasing subsequences are changed to decreasing subsequences and *vice versa*. In particular, $d(\bar{w}) = i(w)$. By Property 1, $d(\bar{w})$ is just the length of the first row of $\bar{P}$. By Property 2, the length of the first row of $\bar{P}$ is just the length of the first *column* of P. Thus $i(w) = \ell(\lambda)$, the number of parts of λ.

We have shown that for a permutation w with $i(w) = p$ and $d(w) = q$, 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 $1, 2, \ldots, n$ with $i(w) = p$ and $d(w) = q$ 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 $i(w) = p$ and $d(w) = q$ 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

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

Hence if $n > pq$, then either $\ell(\lambda) \geq p+1$ or $\lambda_1 \geq q+1$. If we apply the Schensted correspondence 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 $\ell(\lambda) \geq p+1$ or $\lambda_1 \geq q+1$, so by Schensted's theorem either $i(w) \geq p+1$ or $d(w) \geq q+1$.

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 $i(w) = p$ and $d(w) = q$ (when $n = pq$) shows that the Erdős-Szekeres theorem is best possible (see equation (21)). 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 $p \leq q$ (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), …, p (p times), $p+1$ (p times), …, q (p times), $q+1$ ($p-1$ times), $q+2$ ($p-2$ times), …, $p+q-1$ (once). We finally obtain the amazing 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 $1, 2, \ldots, 24$.

6 Reduced decompositions.

There is a remarkable and unexpected connection between standard Young tableaux and the building up of a permutation by interchanging (transposing) two adjacent entries. 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:

$$1\mathbf{23}45 \to 13\mathbf{24}5 \to 1\mathbf{34}25 \to \mathbf{14}325 \to 413\mathbf{25} \to 41352. \tag{22}$$

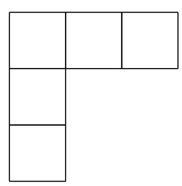
Such sequences of interchanges are used in some of the *sorting algorithms* studied in computer science (see Section 11), although there it is natural to consider the reverse process whereby a list of numbers such as 41352 is step-by-step converted to the "sorted" list 12345. Note that the five steps in the sequence (22) 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 (22) 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 (based on a variant of the Schensted correspondence) 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.

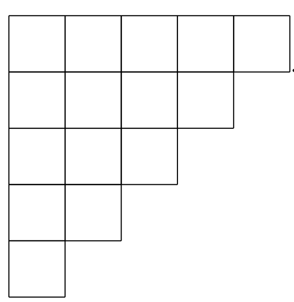
Example. Here are a few examples of the collection $Y(w)$.

(a) If $w = 41352$ (the example considered in equation (22)), 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 (17) or by direct enumeration), it follows that there are six reduced decompositions of 41352.

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



.

Hence

$$r(w) \ = \ f^{(5,4,3,2,1)}$$

$$= \frac{15!}{1^5 \cdot 3^4 \cdot 5^3 \cdot 7^2 \cdot 9}$$

$$= 292,864.$$

(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 (17). 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 reduced decompositions and certain objects in algebraic geometry known as *flag varieties*.

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 for which $i < j$ and $w_i > w_j$, then the parts of λ are just the nonzero α_i's.*

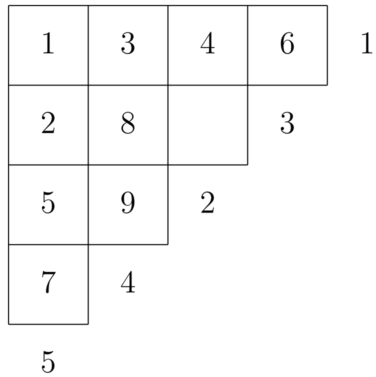
As an illustration of the above theorem, let $w = 526314$. One sees by inspection that w satisfies the conditions of the theorem. We have $(\alpha_1, \ldots, \alpha_6) = (4, 1, 3, 1, 0, 0)$. Hence $\lambda = (4, 3, 1, 1)$ and $r(w) = f^{(4,3,1,1)} = 216$.

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 $1, 2, \ldots, n$ 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 (b. 1964) and Amitai Regev (b. 1940).

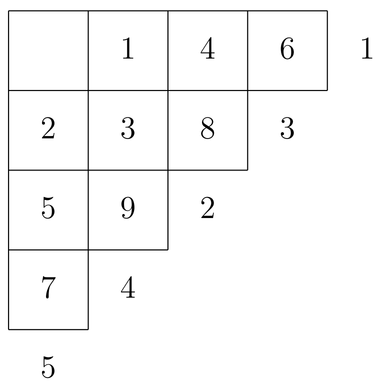
There is one class of vexillary permutations of particular interest. These are the permutations $w_0 = n, n-1, \ldots, 1$, for which $\lambda = (n-1, n-2, \ldots, 1)$. There is an elegant bijection between the SYT of shape $(n-1, n-2, \ldots, 1)$ and the reduced decompositions of w_0, due to Paul Henry Edelman (b. 1956) and Curtis Greene (b. 1944). 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:



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:

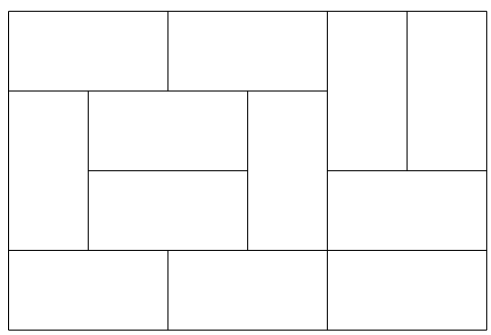
	1	4	6	1
	3	8	3	
2	5	4		
7	2			
5				

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 $12345 \to 13245 \to 13425 \to 14325 \to 14352 \to 41352 \to 41532 \to 45132 \to 45312 \to 45321 \to 54321$.

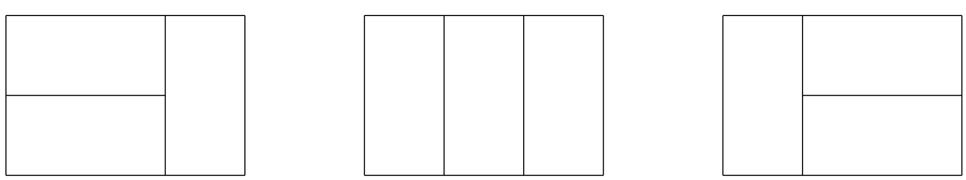
7 Tilings.

The final 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.

The first significant result about the enumeration of tilings was due to the Dutch?? physicist P. W. Kasteleyn (1???–19??) and independently to the British physicist Harold Neville Vazeille Temperley (b. 1931) and the British-born physicist Michael Ellis Fisher (b. 1931). Motivated by work related to the adsorption of diatomic molecules on a surface and other physical problems, they were 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

$$N(2, n) = F_{n+1}, \tag{23}$$

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

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

To prove equation (23), 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 (23).

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 Box), 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, $N(8,8) = 12,988,816 = 3604^2$, 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 Jockusch (19??–) found a direct combinatorial interpretation of $M(2n)$. In 1996 Mihai Adrian Ciucu (b. 1968) 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 5.

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

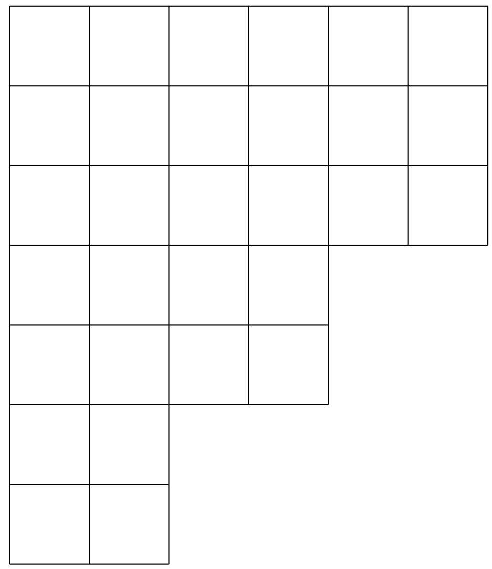


Figure 5: The board R_4.

states that

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

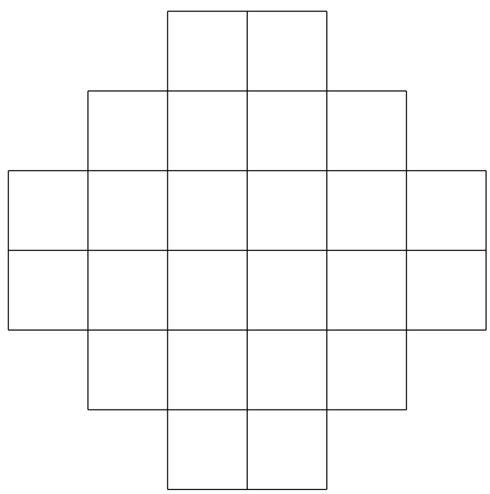
If n is even, say $n = 2r$, then $N(2n,2n) = (2^r T(n))^2$, while if n is odd, say $n = 2r+1$, then $N(2n,2n) = 2(2^r T(n))^2$, so we recover the result that $N(2n,2n)$ is a square or twice a square depending on whether n is even or odd.

BOX. Kasteleyn's formula for the number $N(2m,2n)$ of domino tilings of a $2m \times 2n$ chessboard:

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

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 (b. 1966), Gregory John Kuperberg (b. 1967), Michael Jeffrey Larsen (b. 1962), and James Gary Propp (b. 1960). 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:



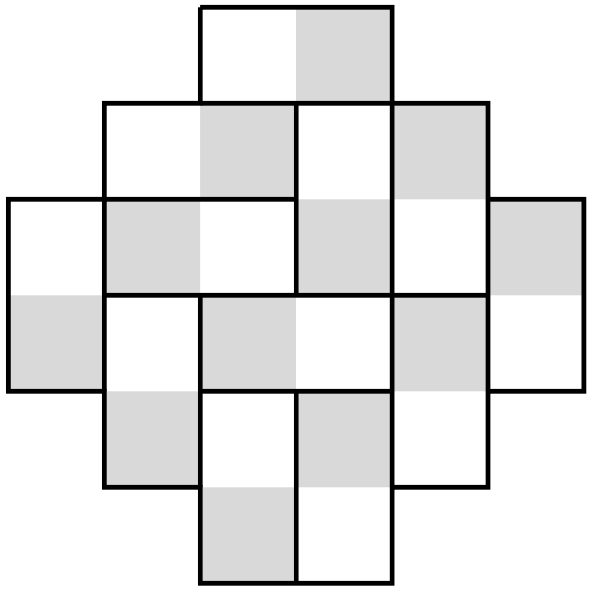
Let $\mathrm{az}(n)$ be the number of domino tilings of the Aztec diamond AZ_n. For instance, AZ_1 is just a 2×2 square, which has two domino tilings (both dominos horizontal or both vertical). Hence $\mathrm{az}(1) = 2$. It's easy to compute by hand that $\mathrm{az}(2) = 8$, and a computer reveals 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

$$\mathrm{az}(n) = 2^{\frac{1}{2}n(n+1)}. \tag{24}$$

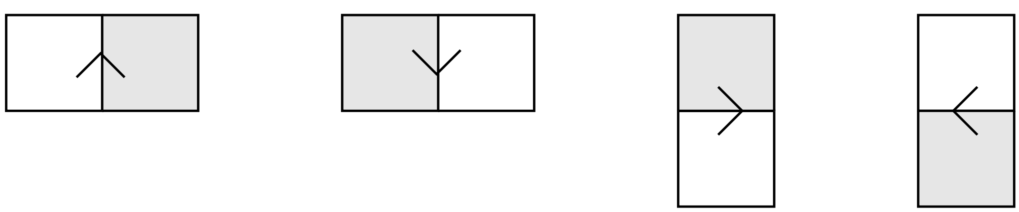
It is rather mysterious why Aztec diamonds seem to be so much more nicely behaved regarding their number of domino tilings than the more natural $m \times n$ chessboards.

A proof of the conjecture (24) 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 $\mathrm{GL}(n+1, \mathbb{C})$.) 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 (24).

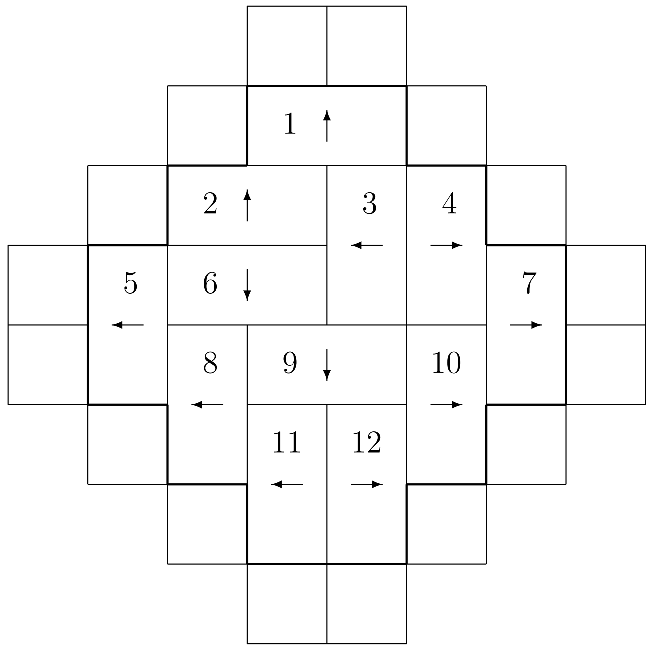
We first color the squares of the Aztec diamond AZ_n black and white in the usual chessboard fashion, with the first (leftmost) square in the top row colored *white*. Here is a tiling of AZ_3 with the chessboard coloring shown.



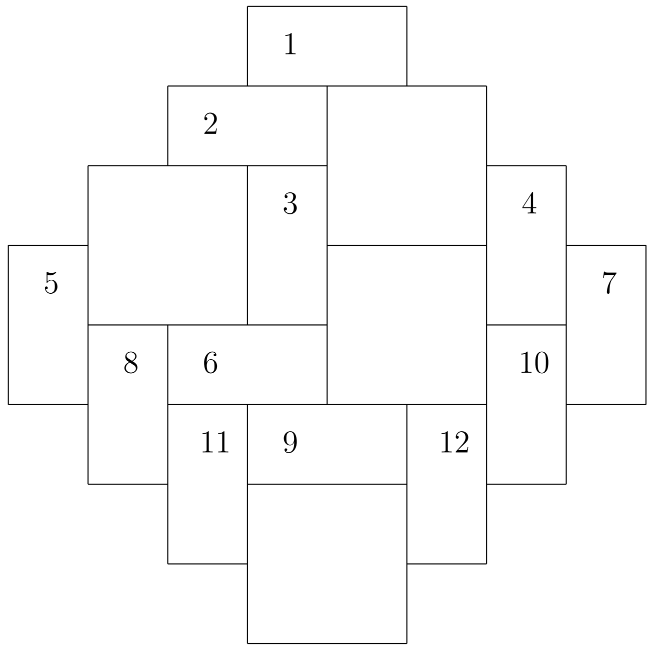
Each domino will have one white square and one black square. There are four possible colorings and orientations of a domino, shown in the illustration below. With each of these four possible colored dominos we associate a direction: up, down, right, and left, as indicated below by an arrow.



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 tiling of AZ_3, with an arrow placed on each domino according to its coloring and orientation, and the boundary of new squares to give AZ_4. We have also numbered each domino for later purposes.



Now move each domino one unit in the direction of its arrow. This is the *shuffling* operation referred to in the name "domino 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 $n+1$ 2×2 squares. The next figure shows the dominos after shuffling (with the same numbers as before), together with the leftover four 2×2 squares.



We now complete the partial tiling of AZ_{n+1} to a complete tiling by putting two dominos in each 2×2 square. There are of course two ways to tile a 2×2 square, so there are 2^{n+1} ways to tile all $n+1$ of the 2×2 squares. Therefore we have associated 2^{n+1} tilings of AZ_{n+1} with each 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 tiling of AZ_n and thus also the $n+1$ 2×2 squares that were left over. Since there are exactly 2^{n+1} tilings of AZ_{n+1} associated with each tiling of AZ_n, 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 (24).

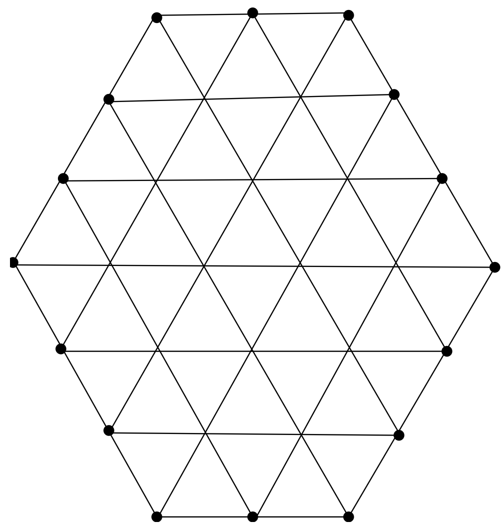


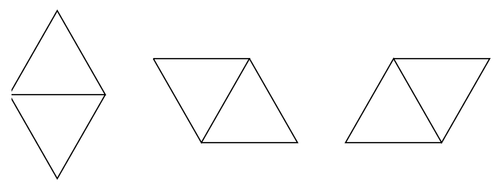
Figure 6: The hexagonal board $H(2, 3, 3)$

8 Tilings and plane partitions.

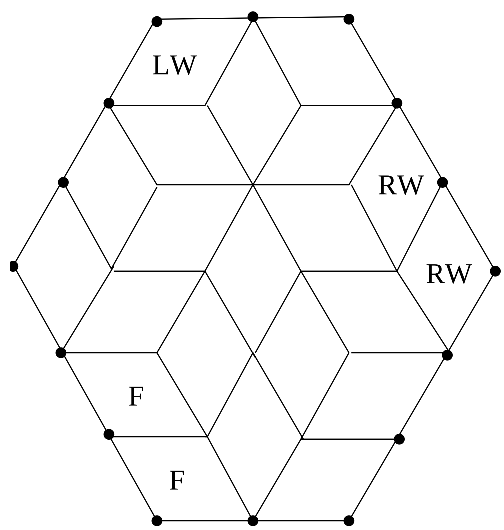
We have discussed several examples of unexpected connections between seemingly unrelated mathematical problems. This is one of the features of mathematics that makes it so appealing to its practitioners. In this section we discuss another such connection, this time between tilings and plane partitions.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

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 6 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.

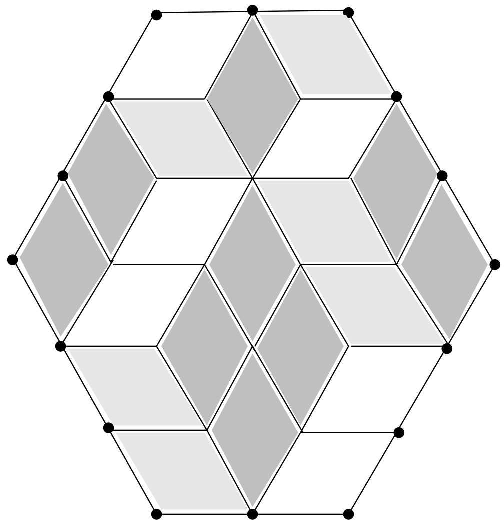
Instead of tiling with dominos (which consist of two adjacent squares), we will be tiling with pieces which consist of two adjacent equilateral triangles. We will call these pieces simply *rhombi*, although they are really only special kinds of rhombi. Thus the number of rhombi in a tiling of $H(r, s,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



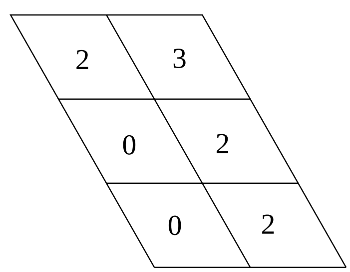
Here is a typical tiling of $H(2, 3, 3)$



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 (14) 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 (14). Therefore substituting $x=1$ is equivalent to replacing $[i]$ by the integer i, so we get the astonishing formula

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

9 Combinatorics and 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 3-space (or, suitably generalized, in higher-dimensional spaces). 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. We will soon exemplify this with two problems coming from computer science. However, first we will discuss in greater detail the connection between topology and combinatorics that will be used.

Let 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 7).

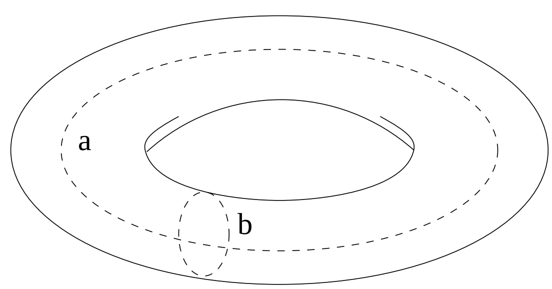


Figure 7: The torus

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 8 shows one way to do 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 8 correspond to the circles marked a and b that are drawn with dashed lines on the torus in Figure 7.

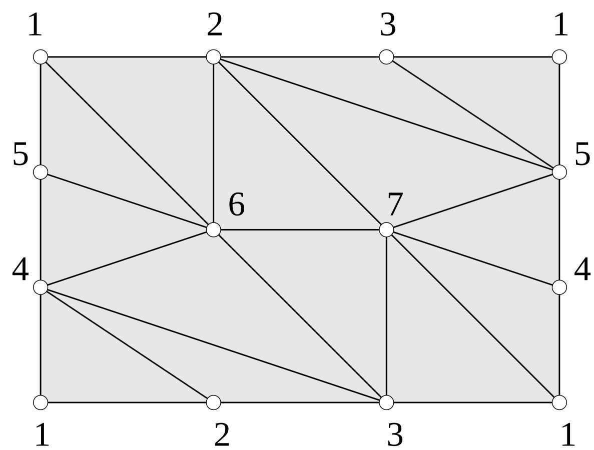


Figure 8: A triangulated torus

Having thus cut the torus apart we now have a collection of 14 triangles. The corners in Figure 8 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

$$\begin{array}{ccccccc} 124 & 126 & 135 & 137 & 147 & 156 & 234 \\ 235 & 257 & 267 & 346 & 367 & 456 & 457 \end{array} \tag{25}$$

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.

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 8 (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!

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 (1854–1912) in the last years of the 1800's and greatly developed in the first half of this 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 measure the number of "independent holes" of various dimensions. It is impossible to give the full technical definition within the framework of this article. 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 very simple (in principle) linear algebra.

For instance, the d-dimensional sphere has Betti numbers $(0, ..., 0, 1)$, reflecting the fact that it has exactly one d-dimensional "hole" (its interior) and no holes of other dimensions. The torus has Betti numbers $(0, 2, 1)$ because there are two essentially different 1-dimensional holes (corresponding to the circles a and b in Figure 7) and one 2-dimensional hole (the interior). Note that the two circles a and b are 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.

The concept of a 0-dimensional hole is perhaps not so clear on an intuitive level, but having $\beta_0 = 0$ means that the space hangs together in one piece (is connected), and in general $\beta_0(T) + 1$ is the number of connected components of the space T. (Note to specialists: Our $\beta_i(T)$'s are really 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.)

We have seen that finite set systems are of use in topology as encodings of topological spaces. But the connection between 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 (25). 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 show that this may indeed be the case, and this is in fact one of the cornerstones for the "topological method" in combinatorics.

The idea to use topological reasoning in combinatorics is quite old but had a somewhat unfortunate start. It seems to have first occurred in connection with a famous problem of Euler. The following configuration is called a *Graeco-Latin square of order* n: An $n \times n$-matrix of ordered pairs (a, b) of numbers a and b from $1, 2, ..., n$ such that the first entries a are distinct in every row and column, the second entries b are distinct in every row and column, and all n^2 possible pairs occur. For instance, here is a Graeco-Latin square of order 3:

$$\begin{array}{ccc} 1,1 & 2,2 & 3,3 \\ 2,3 & 3,1 & 1,2 \\ 3,2 & 1,3 & 2,1. \end{array}$$

Euler stated without proof in his paper "Recherches sur une espèce de carrés magique" from 1782 that such configurations cannot exist for $n = 6, 10, 14, 18, ...$ His claim was proven correct for $n = 6$ by G. Tarry (18??–19??) in 1901. In 1922 Harris F. MacNeish (18??–19??) published a paper in *Annals of Mathematics* supposedly proving Euler's claim for all remaining values of n. His argument, which was based on topology, was unfortunately incorrect. In fact, subsequent research has shown that Euler's claim itself is false, except for the single case of $n = 6$!

After this unsuccessful start it took a long time before the idea resurfaced — topological proofs for combinatorial results have come to the fore only in the last two decades. Let us now go on to see a couple of concrete examples.

BOX: Borsuk and combinatorics

The Polish mathematician Karol Borsuk (1905–1982) 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. Let us state them (within this box we will assume familiarity with the topological terminology used).

Borsuk's Theorem. *If the k-dimensional sphere is covered by $k+1$ closed sets, then one of these sets must contain a pair of antipodal points.*

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

This work of Borsuk has interacted with combinatorics in a remarkable way. In 1978 László Lovász (b. 1948) solved a difficult combinatorial problem — the "Kneser Conjecture" from 1955 — by using Borsuk's theorem. Then, in 1992 the debt to topology was repaid when Jeffry Ned Kahn (b. 1950) and Gil Kalai (b. 1955) solved Borsuk's problem using some results from pure combinatorics. By stating the relevant results on the combinatorial side we hope to give a small glimpse of these interactions, which are quite unexpected.

The answer to Borsuk's problem is definitely "yes" when $k = 1$, the statement then comes down to dividing a line segment of length 1 into two shorter segments, which is clearly possible. 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 that the answer to Borsuk's problem is actually "no", contrary to what "everyone" had believed for nearly 60 years. But one has to go to very high dimensions ($k \approx 1,000$) to find counterexamples with the Kahn-Kalai method. The problem is still open for $k = 4$.

The combinatorial result from which the solution to Borsuk's problem follows is this 1981 theorem of Peter Frankl (b. 19??) and Richard Michael Wilson (b. 1945).

Frankl-Wilson Theorem. *Let k be a power of a prime number, and let F be a family of $2k$-element subsets of $\{1,2,\ldots,4k\}$ such that no two members* *of F have k elements in common. Then F has at most $2 \cdot \binom{4k-1}{k-1}$ members.*

The Kneser conjecture — now a theorem of Lovász — is the following statement:

Lovász' Theorem. *If the n-element subsets of a $(2n+k)$-element set are partitioned into $k+1$ classes, then some class will contain a pair of disjoint n-element sets.*

The details of how this conclusion is derived from Borsuk's theorem, as well as the argument for solving Borsuk's problem using the Frankl-Wilson theorem, must unfortunately be left aside. See the suggested reading for further information.

10 Complexity of graph properties.

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. By constructing algorithms one shows that a task can be done in a certain number of steps. It is often the more difficult part to show that there is no "faster" way, i.e. requiring fewer steps.

Examples of this will be given in this and the following section. We begin by considering 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 9 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.

Connectedness is a very basic property of graphs which can be decided

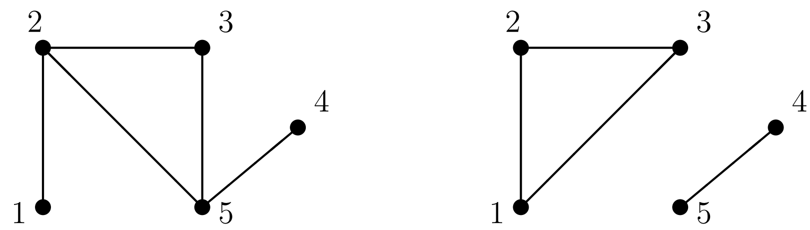


Figure 9: A connected and a disconnected graph

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 this: For a given property P of graphs, is there some algorithm that decides for every graph G whether it has property P without knowing for every pair of nodes whether they span an edge of G or not? If this is not the case, i.e. 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 instance, connectedness is an evasive property. To see this 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 9:

$$\begin{array}{ccccc} * & 1 & 0 & 0 & 0 \\ & * & 1 & 0 & 1 \\ & & * & 0 & 1 \\ & & & * & 1 \\ & & & & * \end{array} \qquad\qquad \begin{array}{ccccc} * & 1 & 1 & 0 & 0 \\ & * & 1 & 0 & 0 \\ & & * & 0 & 0 \\ & & & * & 1 \\ & & & & * \end{array}$$

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: 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 now outline a proof for readers who are developing a taste for combinatorial reasoning, and who understand what is meant by a proof by finite induction.

The crucial step will be to prove the following statement: *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, and let* B *be the set of nodes connected to* j *by* 1*-marked edges. Then all possible edges between nodes in* $A \cup B$ *have been inspected at that stage.* 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 $|A \cup B|$, that is, the number of elements of $A \cup B$. Suppose that $|A \cup B| > 2$. Since 1 (and not 0) is written into position

$\{i, j\}$ 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 $\{c, d\}$ with $c \in C$, $d \in D$ and $\{c, d\} \neq \{i, j\}$ 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.

It has been decided for many graph properties whether they are evasive. It turns out that among the evasive ones are many that 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 completely trivial to verify whereas evasiveness is not, this conjecture — if true — would simplify deciding evasiveness for many graph properties.

The best general result known to date on this topic is the following theorem of Jeffry Kahn, Michael Ezra Saks (b. 1956) and Dean Grant Sturtevant

(b. 1955) 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 $6, 10, 12, 14, 15, ...$; 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.

We will keep in mind some particular monotone graph property P and consider graphs on the nodes $1, 2, ..., n$. 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 10.

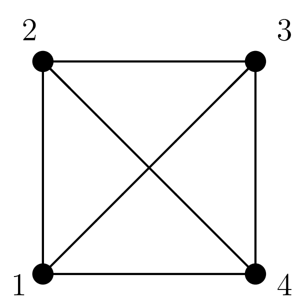


Figure 10: The 6 edges spanned by 4 nodes

The simplicial complex Δ_4^{conn} of disconnected graphs on four vertices is shown in Figure 11.

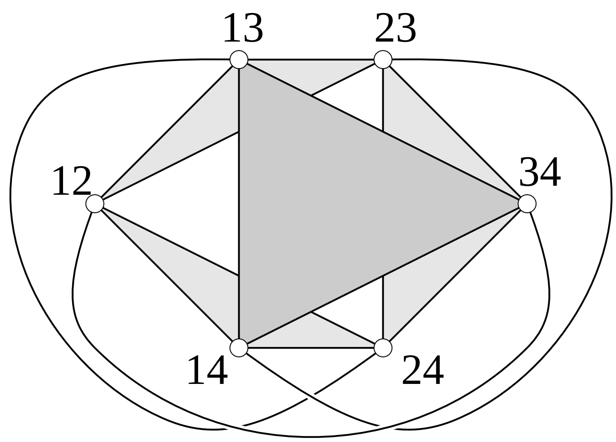
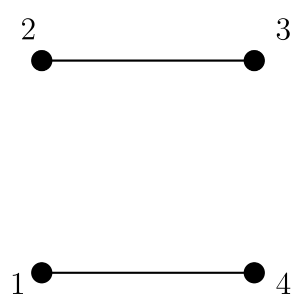
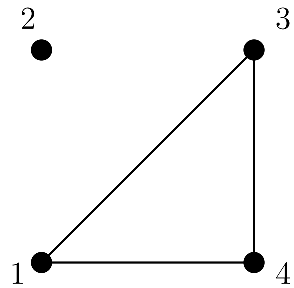


Figure 11: The complex of disconnected graphs on 4 nodes

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 11 corresponds to the disconnected graph



and the triangle with vertices 13, 14 and 34 corresponds to the disconnected graph



Observe in Figure 11 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 (1881–1966) 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 (b. 19??) 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 P *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 (b. 1956) on knot invariants. Also some other monotone graph properties have naturally presented themselves as simplicial complexes in other mathematical contexts.

11 Complexity of sorting and distinctness

The following is a very basic situation studied in complexity theory. A string 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 make 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 $x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_n}$ 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 $2n$ comparisons are needed and that $3n$ 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, which was only recently resolved, is a generalization of the distinctness problem. Namely,

k-equal problem: *for $k \geq 2$, 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 numbers?

247913746858487139551967423461594633148677295592436285411783697 2581932

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 (b. 1947), László Lovász and Andrew Chi-Chih Yao (b. 1946).

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.

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 $\mathbb{R}^n$, 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 $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ 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 $\mathbb{R}^n$. For instance, removal of a plane (a 2-dimensional subspace) cuts $\mathbb{R}^3$ 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 Section 9). 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 $\mathbb{R}^n$ 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 $\mathbb{R}^n$ 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 $\mathbf{x}$ lies on 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 (b. 1950) and Robert Duncan MacPherson (b. 1944), which expresses these Betti numbers in terms of some finite simplicial complexes associated to certain partitions. To get further we need to introduce a few more concepts from combinatorics.

We began this paper by discussing partitions of numbers, and we shall return once more to the ubiquitous concept of partitions. Here we need, however, the notion of *partitions of sets*. A partition of a finite set A is a way of breaking it into smaller pieces, namely a collection of pairwise disjoint subsets whose union is A. (None of these subsets is allowed to be empty — in other words, all the subsets have at least one element.) 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 will use $\{1, 2, \ldots, n\}$ as the ground set and for fixed k (an integer between 2 and n) consider the collection of all partitions of this set 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

There is a natural way to compare set partitions, saying that partition π is less than partition σ (written $\pi \leq \sigma$) 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 12 shows the order diagram of $\Pi_{4,2}$ and Figure 13 shows that of $\Pi_{4,3}$.

These diagrams are 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.

Now, consider the *Möbius function* (see **BOX**) computed over the poset $\Pi_{n,k}$. Let $\mu_{n,k}$ denote the value that the Möbius function attains at the partition with only one part, which is at the top of the order diagram. For example, computation as demonstrated in the **BOX** over the posets in Figures 12 and 13 shows that $\mu_{4,3} = 3$ and $\mu_{4,2} = -6$.

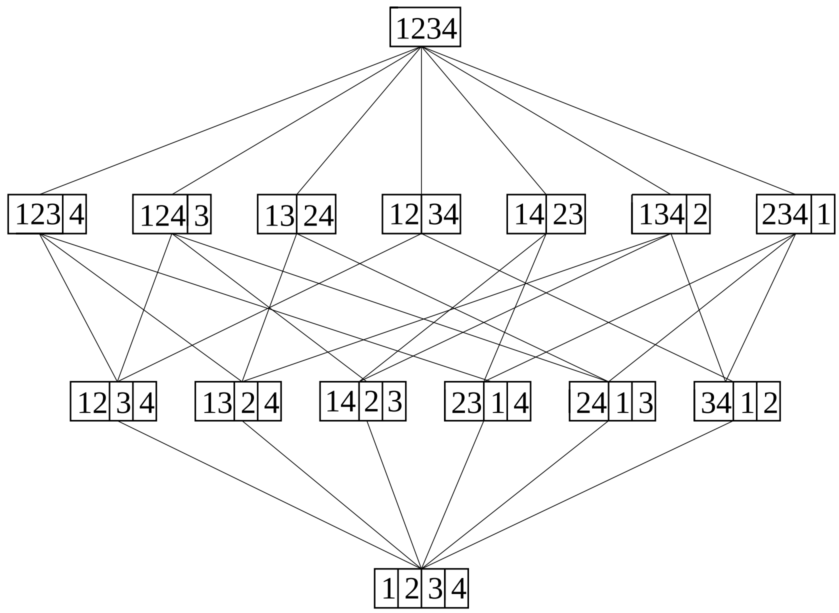


Figure 12: $\Pi_{4,2}$

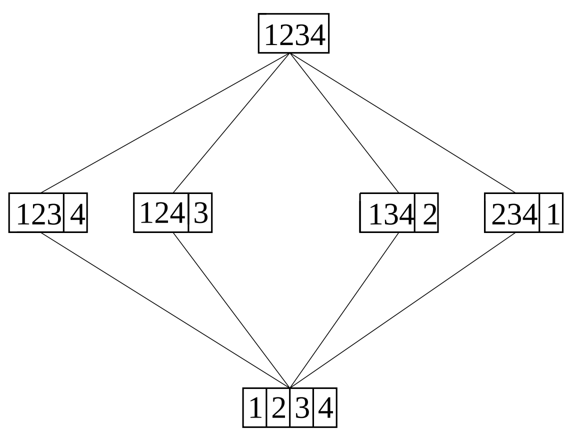


Figure 13: $\Pi_{4,3}$

We can now return to the discussion of the k-equal problem. Where we left off was 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, the following relation:

Fact 2. $\beta(M_{n,k}) \geq |\mu_{n,k}|$.

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 numbers $|\mu_{n,k}|$ grow sufficiently fast. For this we turn to the method of generating functions, already introduced in the early sections 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:

$$\exp\left(\sum_{n\geq 1} \mu_{n,k}\frac{x^n}{n!}\right) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{k-1}}{(k-1)!}. \tag{26}$$

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

$$\mu_{n,2} = (-1)^{n-1}(n-1)!.$$

Also, in this case the formula (26) 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

$$\mu_{n,3} = -(n-1)!\,\big((-1+i)^{-n} + (-1-i)^{-n}\big) = -(n-1)!\,2^{1-\frac{n}{2}}\cos\frac{3\pi n}{4}. \quad (27)$$

We have come to a point where we know on the one hand from Facts 1 and 2 that

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 (27) 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.

BOX: The Möbius function.

The *Möbius function* is one of the most important tools of algebraic combinatorics. It assigns a very significant integer to every finite "poset". This word is an abbreviation which stands for "partially ordered set"; for simplicity we will assume that all posets considered have a bottom and a top element. Figure 14 shows a poset of eight elements with bottom element "a" and top element "h".

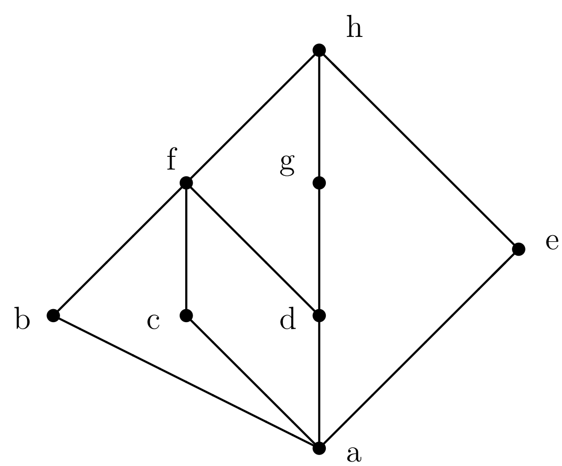


Figure 14: A small poset

The Möbius function $\mu(x)$ is recursively defined for any finite poset as follows: Put $\mu(x_0) = 1$ for the bottom element x_0 of the poset, then require that

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

for all other elements x. This formula means that we are to define $\mu(x)$ so that when we sum $\mu(y)$ for all y less than or equal to x the resulting sum equals zero. This can clearly be done as long as one knows the values $\mu(y)$ for all elements y less than x. The reader can see how this recursive definition works by computing the Möbius function of the poset in Figure 14, 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 15 shows the same poset with computed Möbius function values.

The Möbius function has its origin in number theory, where it was introduced by August Ferdinand Möbius (1790–1868). (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 instance, see the divisor diagram of the number "60" in Figure 16. A calculation based on this diagram, analogous to the one we just carried out over Figure 14, will show 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 prime factors.

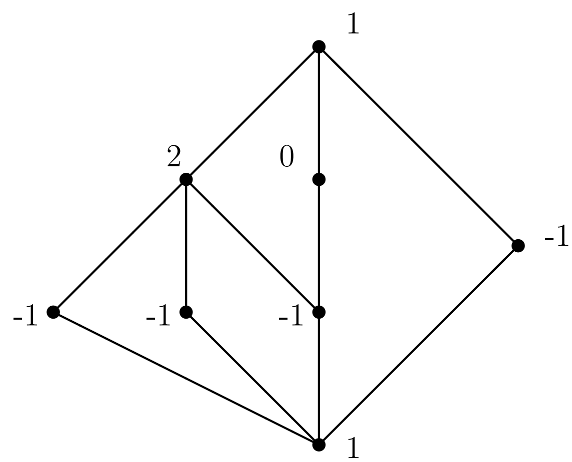


Figure 15: Values of the Möbius function

The Möbius function is very important 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

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

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

The Möbius function is an indispensable tool in enumerative combinatorics because 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 $f : P \to \mathbb{Z}$ from a poset P to the integers is related to another function $g : P \to \mathbb{Z}$ by the partial summation formula

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

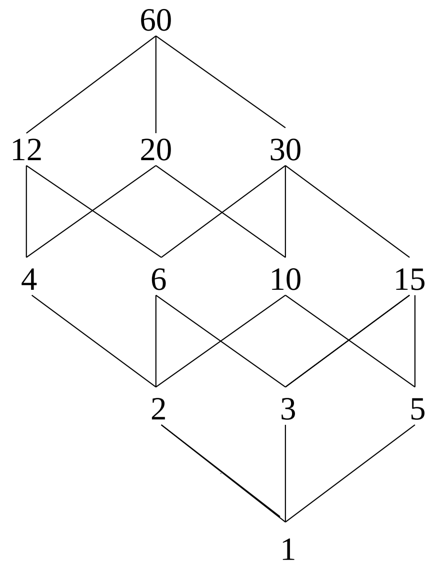


Figure 16: The divisors of “60”.

then the value $g(x_0)$ at the bottom element x_0 of P can be expressed in terms of f via the formula

$$g(x_0) = \sum_{y \in P} \mu(y) f(y).$$

The Möbius function also has a topological meaning, which is the reason it turns up in “Fact 2” of this section. The connection is as follows. Let P be a poset with bottom element b and top element t. Define the set family $\Delta(P)$ to consist of all chains (meaning: totally ordered subsets) $x_1 < x_2 < \cdots < x_k$ in $\overline{P} = P \backslash \{b, t\}$, meaning P with b and t removed. Then $\Delta(P)$ is a simplicial complex (since a subset of a chain is also a chain), so as discussed in Section 9 there is an associated topological space.

For instance, let P be the divisor poset of the number “60” shown in Figure 16. Then $\overline{P} = P \setminus \{1, 60\}$ has the following twelve maximal chains

2 — 4 — 12
2 — 4 — 20

2 — 6 — 12
2 — 6 — 30
2 — 10 — 20
2 — 10 — 30
3 — 6 — 12
3 — 6 — 30
3 — 15 — 30
5 — 10 — 20
5 — 10 — 30
5 — 15 — 30

As was explained in Section 9 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 17 — a 2-dimensional disc.

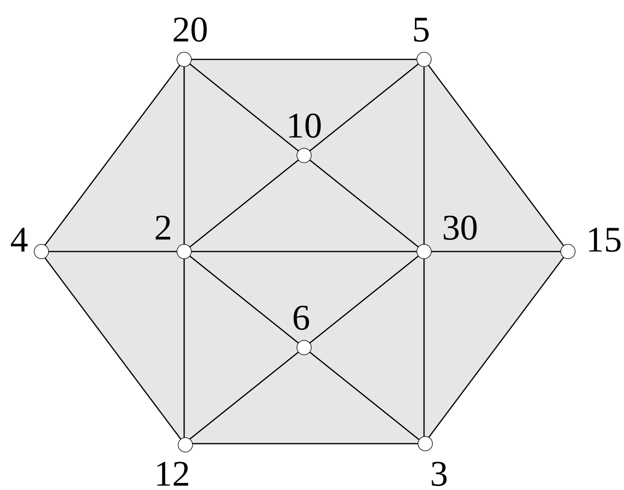


Figure 17: The simplicial complex of proper divisors of "60".

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,

$$\mu(P) = \beta_0(P) - \beta_1(P) + \beta_2(P) - \beta_3(P) + \cdots . \tag{28}$$

For instance, the space depicted in Figure 17 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 (28) that $\mu(P) = 0$. This "explains" topologically why $\mu(60) = 0$, a fact we already knew from simpler considerations. On the other hand, if P is the divisor diagram of "30" (which can be seen as a substructure in Figure 16), then $\Delta(P)$ is the circle 2 — 6 — 3 — 15 — 5 — 10 — 2 (a substructure in Figure 17). This circle has a one-dimensional hole, so $\beta_1(P) = 1$. All other Betti numbers are zero, hence formula (28) gives that $\mu(30) = -1$, another fact we have already encountered.

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 "Euler's formula" for 3-dimensional polytopes from 1752. It goes as follows.

A 3-*polytope* P (or, 3-dimensional convex polytope, to be more precise) is for a mathematician a bounded region of space obtained as the intersection of finitely many halfspaces (and not contained in any plane). 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. For instance, take the ordinary cube shown in Figure 18 — 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 polytope that is dear to all combinatorialists is the "permutohedron", shown in Figure 19. 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 permu-

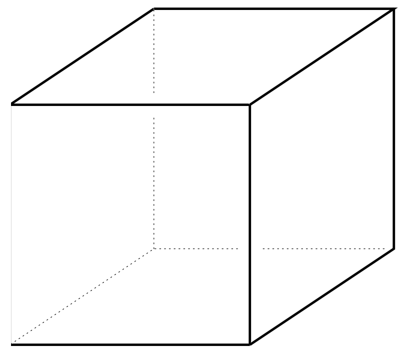


Figure 18: The cube.

tations 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 Section 6.

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, see Figures 18 and 19.

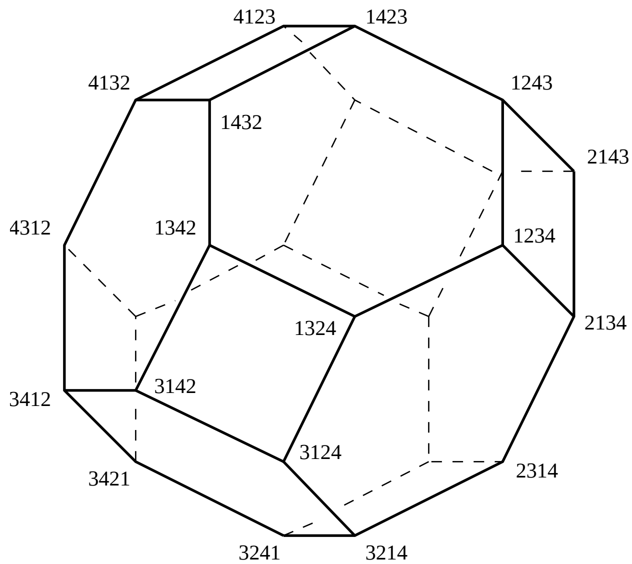


Figure 19: The permutohedron.

	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 $\mathbb{R}^d$. Such higher-dimensional polytopes have taken on great practical significance in the last fifty years because of 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 plane surface in $\mathbb{R}^d$"), 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:*

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

However, 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 Henri Poincaré.

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

$$(14, 89, 338, 850, 1484, 1834, 1604, 971, 380, 76) \tag{29}$$

is the f-vector of a 10-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

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

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 will now formulate.

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 20. In general, a d-simplex is the natural d-dimensional analogue of the tetrahedron.

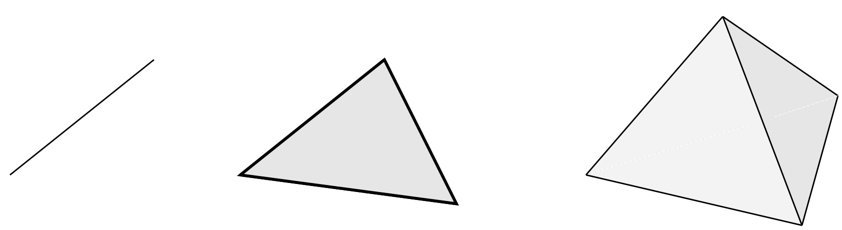


Figure 20: A d-simplex, $d = 1, 2, 3$.

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 21; so the octahedron and icosahedron are examples of simplicial polytopes

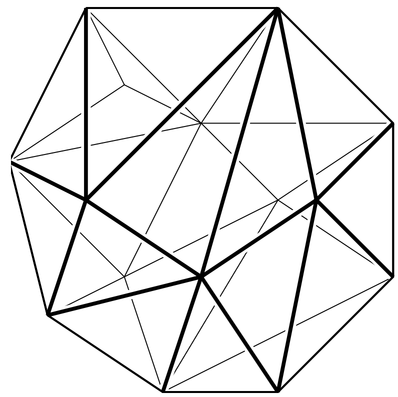


Figure 21: A simplicial 3-polytope.

but the cube and permutohedron are not. If a polytope is simplicial then its faces form a simplicial complex in the sense defined in Section 9. 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.

In 1970 Peter McMullen (b. 1942) 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 (b. 1943) and Carl William Lee (b. 1954) and one by Richard Peter Stanley (b. 1944), 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 M of monomials in indeterminates $x_1, x_2, \ldots, x_n$ such that if $m \in M$ and m' divides m then $m' \in M$. Figure 22 shows the multicomplex $M = \{1, x, y, z, x^2, xy, yz, z^2, x^2y, z^3\}$ ordered by divisibility.

An M-*sequence* is a sequence $(1, a_1, a_2, a_3, \ldots)$ such that each a_i is the

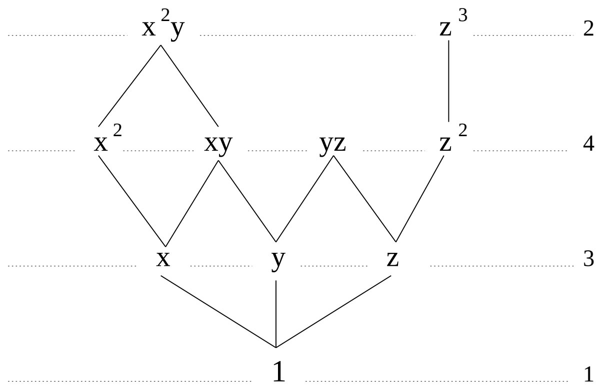


Figure 22: A multicomplex.

number of monomials of degree i in some fixed multicomplex. For instance, the M-sequence coming from the multicomplex M in Figure 22 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 (1862-1937) who first seems to have 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.

We are now ready to formulate the g-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

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

Here we are using the *binomial coefficients*, defined by

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

where $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$, and $0! = 1$. (The factorial $n!$ was already used in connection with equation (17), but we repeat the definition here for the reader's convenience.)

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*

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

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

The equation $\mathbf{f} = \mathbf{g} \cdot M_d$ 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 $\mathbf{f}$ displayed in equation (29) 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

$$\mathbf{f} = (1, 3, 4, 2, 0, 0) \cdot M_d,$$

and we know from Figure 22 that $(1, 3, 4, 2, 0, 0)$ is an M-sequence. Hence, $\mathbf{f}$ 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. We will however not state it here.

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 $\mathbf{g}$ 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 paper 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. Take for example the Schensted correspondence described in Section 3, connecting permutations and pairs of standard Young tableaux; or the connections between combinatorics and representation theory or combinatorics and topology described in earlier sections. The g-theorem is one more example of this kind, establishing an 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. Thus, the study of f-vectors, initiated by Euler's discovery almost 250 years ago, is likely to remain an important challenge for many years to come.

13 Further reading (incomplete)

We refer here mainly to general accounts that should be at least partially accessible to the layman and that give lots of further references.

For a broad view of current combinatorics, with a wealth of information and references, see

- *Handbook of Combinatorics* (eds. R. Graham, M. Grötschel and L. Lovász), North-Holland, Amsterdam, 1995.

A good reference for number partitions is

look for popular articles

- G.E. Andrews, *Theory of partitions*, ...

The basic theory of enumeration is developed in

- R.P. Stanley, *Enumerative Combinatorics*, Volume 1 (Wadsworth & Brooks/Cole, Monterey, CA, 1986) and Volume 2 (Cambridge University Press, Cambridge, UK, to appear in 1997/8?).

The combinatorics of number and set partitions, standard Young tableaux, generating functions and the Möbius function, together with algebraic ramifications, is discussed there. A briefer account of this material is given in

- I. Gessel and R.P. Stanley, *Algebraic Enumeration*, a chapter in the "Handbook of Combinatorics", pp. xxx–yyy

A wealth of information about the topic of tilings can be found in

- Grünbaum and Shepard

For enumerative aspects of tilings see

- Elkies-Kuperberg-Larsen-Propp paper in J. Alg. Comb.

The following book is a nice companion to the study of enumeration

- N. J. A. Sloane and S. Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, 1995. There is also an interactive version on the web at http://www.research.att.com/∼njas/sequences/

For connections between combinatorics and topology, including more details about the evasiveness and Kneser conjectures, see either of

- A. Björner, *Combinatorics and Topology*, Notices of the American Mathematical Society **32** (1985), 339–345.
- A. Björner, *Topological Methods*, a chapter in the "Handbook of Combinatorics", pp. 1819–1872.

The disproof of the Borsuk conjecture is reported in

- B. Cipra, *Disproving the obvious in higher dimensions*, What's Happening in the Mathematical Sciences **1** (1993), 21–25,
- A. Skopenkov, *Borsuk's problem,* Quantum **7** (1996), 17–21,

while more about the k-equal problem and its solution can be found in

- A. Björner, *Subspace Arrangements*, in "First European Congress of Mathematics, Paris 1992" (eds. A. Joseph *et al*), Progress in Mathematics Series, Volume **119**, Birkhäuser, Boston, 1994, pp. 321–370.

Finally, for convex polytopes and the g-theorem we refer to

- G. M. Ziegler, *Lectures on Polytopes*, GTM Series, Springer-Verlag, Berlin, 1995.