A SURVEY OF ENUMERATIVE COMBINATORICS

Lectures by Alex Postnikov
Notes by Travis Dillon

Contents

The goal of this course is to survey enumerative combinatorics, largely pulling from *Enumerative Combinatorics, Volumes I and II*.

1. the twelvefold way

1.1. introduction

The twelvefold way is a set of twelve counting problems that introduce basic notions in combinatorics, conveniently bundled into a single table. In Stanley's book, they're all phrased in terms of counting the number of functions between two finite sets subject to certain restrictions. But this terminology is a bit awkward (as evidenced by his switch midway through his discussion of the way to a metaphor of balls and boxes).

So here's how we'll imagine it. We have a collection M of m marbles and a collection B of b bags. These collections are not necessarily sets; sometimes, we may consider the elements of M or N indistinguishable. Think of this as having a collection of identical marbles or bags: You can ascertain the groupings, but you can't tell one individual marble (or bag) from the other.

We want to count the number of ways to place marbles into bags in ways that satisfy certain restrictions. If we consider a placement of marbles into bags as a function $M \to B$, the restrictions are that this function is injective, surjective, or generic (neither). These restrictions correspond to counting the number of ways of placing marbles into bags such that no bag has more than one ball; every bag has at least one ball; and no restrictions. In all, we get this table:

The twelvefold table

M	B	—	inj.	surj.
dist.	dist.	(1)	(2)	(3)
ind.	dist.	(4)	(5)	(6)
dist.	ind.	(7)	(8)	(9)
ind.	ind.	(10)	(11)	(12)

We'll deal with the table one entry at a time.

1.2. the easy entries

Entry (1). For each marble, we have a choice of b different bags to place it in, so there are b^m distributions.

Entry (2). There are b possible bags to place the first marble in, $b-1$ bags for the second, and so on, so there are $b(b-1)\cdots(b-m+1)$ distributions. This number is called the mth *falling power* of b and is denoted $b^{\underline{m}}$.[1]

Entry (5). A distribution of marbles corresponds to choosing m different bags out of the b total, so there are $\binom{b}{m}$ of them.

Entry (8). There is one distribution if $m \le b$—each marble is placed in one of the indistinguishable bags—and no distribution if $m \ge b$.

Entry (11). Same as entry (8).

Entries (8) and (11) can be succinctly formulated using the delta function, which takes in a proposition P and returns 1 if P is true and 0 if P is false. The values of (8) and (11) are $\delta(m \le b)$.

[1] This number is also called the mth *falling factorial* of b; it is sometimes denoted $(b)_m$, which is (to me) less intuitive. The mth *rising power* or *rising factorial* of b, which is defined $b^{\overline{m}} = b(b+1) \quad (b+m \quad 1)$; in the alternate notation, this is denoted $b^{(m)}$.

1.3. stars and bars and donuts and dividers

Entry (4). We can imagine placing the bags in order, taking out their marbles, and placing them in a line with a vertical divider between the marbles from consecutive bags. If the first bag has 3 marbles, the second 2, and the third 4, then the diagram we get would look like this:

$$\bullet\bullet\bullet \mid \bullet\bullet \mid \bullet\bullet\bullet\bullet$$

Counting the number of these diagrams is not so hard: There are $m+b-1$ positions; in m positions, we place balls, and in the other $b-1$ we place dividers. So there are $\binom{m+b-1}{m} = \binom{m+b-1}{b-1}$ different distributions.

Entry (6). This is the same as for entry (4), except we need one ball in each section. Notice, though, that if we subtract one ball from each section, we get exactly the set of diagrams coming from $m-b$ balls and b bags. So there are $\binom{m-1}{m-b} = \binom{m-1}{b-1}$ different distributions.

An alternate explanation: To get a diagram where each section contains at least one ball, we need only choose to place $b-1$ dividers from the $m-1$ different spaces between the line of m balls. This, again, is $\binom{m-1}{b-1}$.

When this concept is introduced in the States, often F is drawn instead of $\bullet$, so this little trick goes by the name of "stars and bars." In some other places, it goes by the alliterative name "donuts and dividers," and I suppose they draw little donuts instead of circles.

These numbers are intimately related to something called a *composition* of a natural number.

Definition 1.1. A *composition* of a positive integer $n \in \mathbb{N}$ into k parts is a sequence $(\lambda_1, \lambda_2, \dots, \lambda_k)$ of positive integers which sum to n.

We think of a composition of n as a way to break n down into the sum of k different integers where *order matters*. For example, the compositions of 5 into 3 parts are

$$\begin{array}{cc} (3,1,1) & (2,2,1) \\ (1,3,1) & (2,1,2) \\ (1,1,3) & (1,2,2). \end{array}$$

Exercise 1.1. Find a formula for the number of compositions of n into k parts.

Exercise 1.2. Show that there are 2^{n-1} different compositions of n into any number of parts.

1.4. stirling numbers

Definition 1.2. The *Stirling numbers of the second kind*, denoted $S(n,k)$, are the number of ways to partition an n-element set into k nonempty subsets.

For example, $S(4,2) = 6$, since there are six different ways to partition $\{1,2,3,4\}$ into two nonempty subsets (four partitions split into a 3-element and a 1-element subset; the other two split into two 2-element subsets). Trivially, $S(n,1) = S(n,n) = 1$ for every n. For simplicity, we introduce the notation $[n] := \{1, 2, \dots, n\}$.

Exercise 1.3. Prove that $S(n, n-1) = \binom{n}{2}$ and $S(n,2) = 2^{n-1} - 1$.

Entry (9). This is $S(m,b)$. It seems cheap to define away the problem, but that's the Way It Is.

Entry (3). This is the same as entry (9), except we order the blocks of the partition. There are $b!$ ways to order b sets, so there are $b!S(m,b)$ distributions.

Entry (7). Now we want to partition M into b different sets, but some of these sets might be empty. There are $S(m, b-k)$ ways to partition where k of the sets are empty, so the total number of distributions is $\sum_{k=1}^{m} S(m,k)$.

Definition 1.3. The nth *Bell number*, denoted B_n, is the number of ways to partition an n-element set into nonempty subsets.

The Stirling numbers have all sorts of interesting behavior. For example, they translate between the falling and regular powers:

Proposition 1.4. *$x^n = \sum_{k=1}^{n} S(n,k)x^{\underline{k}}$ for all $n \in \mathbb{N}$.*

Proof. Fix some $n \in \mathbb{N}$. To show this equality, it suffices to check it for $n+1$ distinct values of x; in particular, if the statement is true on the positive integers, then it is true for all real values of x.[2] To prove this, we claim that both sides count the number of functions $[n] \to [x]$. Certainly the left one does; how does the right? We first group the functions $[n] \to [x]$ by the size of their image. If $|\mathrm{im}(f)| = k$, then the preimages of each element form a partition of $[n]$ into k nonempty sets; for each partition of $[n]$, there are $x^{\underline{n}}$ maps that send each partition to a single, unique element of $[x]$. So there are $S(n,k)x^{\underline{n}}$ functions whose image has size k. Sum over all values of k to complete the proof. □

A more abstract way of looking at this is to consider the vector space $P_n(\mathbb{R})$, the set of real polynomials whose degree is at most n. The set $\{1, x, x^2, \ldots, x^n\}$ is a basis for $P_n(\mathbb{R})$, but so is $\{1, x, x^{\underline{2}}, \ldots, x^{\underline{n}}\}$. The Stirling numbers of the second kind provide a change-of-basis matrix $\mathbf{S} = \big(S(n,k)\big)_{n,k}$:

$$\begin{pmatrix} S(n,n) & 0 & 0 & \cdots & 0 \\ S(n,n-1) & S(n-1,n-1) & 0 & \cdots & 0 \\ S(n,n-2) & S(n-1,n-2) & S(n-2,n-2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S(n,0) & S(n-1,0) & S(n-2,0) & \cdots & S(0,0) \end{pmatrix}$$

where $S(n,0) = 0$ if $n < 0$ and 1 if $n = 0$. If $\mathbf{Sa} = \mathbf{b}$, then $\sum_{k=0}^{n} a_k x^k = \sum_{k=0}^{n} b_k x^{\underline{k}}$.

Finally, we note a formula for the Stirling numbers:

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

This can be proven by multiplying both sides by $k!$ and then using Inclusion-Exclusion to show that the right side counts the number of surjective functions $[n] \to [k]$. (We do this in Section 1.5.)

So we didn't cheat too much when we defined the problem away.

1.5. an aside: inclusion-exclusion

The Principle of Inclusion-Exclusion can be stated like this:

Theorem 1.5. *If $A_1, \ldots, A_n$ are finite sets, then*

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{\emptyset \neq S \subseteq [n]} (-1)^{|S|-1}\left|\bigcap_{j \in S} A_j\right|.$$

This formulation is useful when you want to count the number of items that satisfy at least one of several different conditions. An equivalent way to state the Principle is useful when you want to count the number of items that satisfy *none* of the conditions:

[2] In fact, if it's true for all $x \in \mathbb{N}$, then it's true as a statement about formal polynomials over $\mathbb{Z}$, but we needn't be that abstract.

Theorem 1.6. *If X is a finite set and $A_1, \ldots, A_n \subseteq X$, then*

$$\left| X \setminus \bigcup_{i=1}^{n} A_i \right| = \sum_{S \subseteq [n]} (-1)^{|S|} \left| \bigcap_{j \in S} A_j \right|,$$

where $\bigcap_{i \in \emptyset} A_i = X$ by convention.

In particular, if $\left|\bigcap_{j \in S} A_j\right| = a_k$ whenever $|S| = k$, the formula simplifies to

$$\left| X \setminus \bigcup_{i=1}^{n} A_i \right| = \sum_{k=0}^{n} (-1)^k \binom{n}{k} a_k. \tag{1.1}$$

How do you prove it? You can use induction, but that gives you no idea what's going on. Here's a different method which is slightly more enlightening. It all comes down to this basic fact:

Lemma 1.7. $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$ *for every* $n \in \mathbb{N}$.

Proof 1. Substitute $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ and expand; everything cancels. □

Proof 2. This is the binomial expansion for $\big(1 + (-1)\big)^n = 0$. □

But a bijective proof is nicer:

Proof 3. We want to show that the number of even subsets of $[n]$ is the same as the number of odd subsets of $[n]$. Taking the symmetric difference of a set with $\{n\}$ is a bijection between these two collections. □

This proof is basically the same as the first: Can you see how?

Anyway, back to Inclusion-Exclusion.

Proof of Theorem 1.6. Pick any element $x \in X$; we show that the number of times it is counted on the left side of the equation is exactly the same as the number of times it is counted on the right side of the equation.[3] If $x \notin \bigcup_{i=1}^{n} A_i$, then x is counted on both sides exactly once. If $x \in \bigcup_{i=1}^{n} A_i$, let s denote the total number of sets A_i which contain x. Then on the right, x is counted

$$\sum_{k=0}^{s} (-1)^k \binom{s}{k} = 0$$

times, as desired. □

To prove the formula for the Stirling numbers of the second kind, let A_i be the set of functions $[n] \to [k]$ whose image does not contain i. Then you can apply (1.1) to get

$$S(n,k) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} (n-i)^n;$$

reversing the order of the dummy variable i gives the stated formula.

[3] If you want, what we are proving is a corresponding statement in the free abelian group generated by X: If we set $\sigma(A) = \sum_{a \in A} a$, then we prove that

$$\sigma(X) - \sigma\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{S \subseteq [n]} (-1)^{|S|} \sigma\Big(\bigcap_{j \in S} A_j\Big).$$

1.6. partitions

Definition 1.8. A *partition* of n into k parts is a sequence $(\lambda_1, \ldots, \lambda_k)$ of weakly decreasing positive integers that sum to n. In other words, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ and $\lambda_1 + \cdots + \lambda_k = n$. The number of partitions of n into k parts is denoted $p_k(n)$, and the number of partitions of n into any number of parts is denoted $p(n)$.

Item (12). This is $p_b(m)$, pure and simple.

Item (10). This is the number of partitions of m into at most b parts. Every partition of $m + b$ into b parts can be made into a partition of m into at most b parts by subtracting 1 from each part. This is a bijection (make sure you see why!), so there are $p_b(m + b)$ total distributions.

Again we've defined away the problem, but this time there's actually no way around it. There is no formula for the partition function, and studying it actually leads you quite deep into analytic number theory. We'll just do a recurrence relation.

Exercise 1.4. Prove that $p_k(n) = p_k(n-k) + p_{k-1}(n-1)$.

1.7. the completed table

The twelvefold table

M	B	—	inj.	surj.
dist.	dist.	b^m	$b^{\underline{m}}$	$b!S(m,b)$
ind.	dist.	$\binom{m+b-1}{m}$	$\binom{b}{m}$	$\binom{m-1}{b-1}$
dist.	ind.	B_n	$\delta(m \leq b)$	$S(m,b)$
ind.	ind.	$p_b(m+b)$	$\delta(m \leq b)$	$p_b(m)$

Robert Proctor extended this table to a "Thirtyfold Way" in a 2006 preprint.

2. generating functions

2.1. introduction

Definition 2.1. The *ordinary generating function* of a sequence $(a_n)_{n=0}^{\infty}$ is the power series $\sum_{k=0}^{\infty} a_k x^k$. The *exponential generating function* of (a_n) is $\sum_{k=0}^{\infty} a_k \frac{x^n}{k!}$. We write $f \leftrightarrow (a_n)$ if f is the ordinary generating function of the sequence (a_n) and $g \xleftrightarrow{\text{exp}} (b_n)$ if g is the exponential generating function of (b_n). The notation $[x^n]f$ means the coefficient of x^n in f (so $[x^n]f = a_n$).

Usually, we consider these not as functions, as the name implies, but as formal power series in the ring $\mathbb{Q}[[x]]$ (or perhaps $\mathbb{R}[[x]]$ or $\mathbb{C}[[x]]$, but $\mathbb{Q}$ is usually all we need).

Herbert Wilf gave a famously good turn of phrase about them:

> A generating function is a clothesline on which we hang up a sequence of numbers for display.

Though I prefer his other quip

> In mathematics it is always best to cheat.

Anyway, the useful part of generating functions is that we can import the algebra of polynomials and power series, and this tells us something about the original sequence. So let's get to that algebra.

Proposition 2.2. *If $f \leftrightarrow (a_n)$ and $g \leftrightarrow (b_n)$, then the elements of the sequence $(c_n) \leftrightarrow f \cdot g$ are*

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

Proposition 2.3. *If $f \xleftrightarrow{\text{exp}} (a_n)$ and $g \xleftrightarrow{\text{exp}} (b_n)$, then the elements of $(c_n) \xleftrightarrow{\text{exp}} f \cdot g$ are*

$$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}.$$

Proposition 2.2 is just regular polynomial or power series multiplication; the second is just a little bit more.

Exercise 2.1. Prove Proposition 2.3.

Here's the combinatorial interpretation of generating function multiplication. Suppose that $f \leftrightarrow (a_n)$ and $g \leftrightarrow (b_n)$, and think of (a_n) as counting the number of A-structures you can place on n unlabelled nodes and (b_n) as the number of B-structures that you can place on n unlabelled nodes. (These A- and B-structures might be graphs, trees, orderings, etc.) The coefficient $[x^n]f \cdot g$ is the number of ways to partition n unlabelled nodes into two sets and place an A-structure on the first and a B-structure on the second. For example, if f is the ordinary generating function for the number of unlabelled trees with n vertices, then f^2 is the generating function for the number of forests with n vertices and exactly two trees.

Example 2.4. Let $f \leftrightarrow (1)_{n=0}^{\infty}$, so $f = 1 + x + x^2 + x^3 + \cdots$. The power series f^k is the generating function for the number of compositions of n into k parts, so its coefficients are

$$[x^n]f^k = \binom{k+n-1}{n}.$$

If $f \xleftrightarrow{\text{exp}} (a_n)$ and $g \xleftrightarrow{\text{exp}} (b_n)$, we can think of a_n as the number of A-structures that can be placed on n *labelled* nodes and b_n the number of B-structures that can be placed on n labelled nodes. Then $f \cdot g$ is the generating function for the number of ways to take n labelled nodes, partition them into two sets, and place an A-structure on the first set and a B-structure on the second. If g is the exponential generating function for the number of labelled trees with n vertices, then g^2 is the number of labelled forests with exactly two trees.

Proposition 2.5. *A power series $f \leftrightarrow (a_n)$ has a multiplicative inverse g (a power series such that $fg = gf = 1$) if and only if $a_0 \in \{\pm 1\}$.*[1]

Example 2.6. The inverse of $1 - kx$ is the power series $1 + kx + (kx)^2 + \cdots$.

Example 2.7. We let $P \leftrightarrow \big(p_k(n)\big)_{n \ge 0}$. We have

$$P(x) = \prod_{k=1}^{\infty} \frac{1}{1-kx}.$$

[1] In fact, this is a special case of a more general theorem that, when A is a commutative ring, an element of $A[[x]]$ is invertible if and only if its constant term is a unit in A.

To see this, expand out each term on the right to get

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

When multiplying out, each monomial results from choosing an element in each series, something like $x^n = x^{i_1}x^{2i_2}x^{3i_3}\cdots$. This corresponds exactly to a partition of n with i_1 ones, i_2 twos, i_3 threes, and so on. So the coefficient of x^n in the power series on the right is $p(n)$,

(This seemingly doesn't account for all terms: What if we choose x in each series? If you wish, you can define the infinite product as the limit of the finite partial products; this gets what is described here.)

Composition is also an interesting thing, but the problem is that it's not always well-defined. For example, let's take $f = 1+x$ and $g \leftrightarrow (1)_{n \geq 0}$. Then

$$g \circ f = 1 + (1+x) + (1+x)^{+}(1+x)^3 + \cdots$$

which makes every coefficient infinite! But if f and g are two power series and the constant term of g is 0, then the coefficient of x^n in $g \circ f$ is always determined by a finite calculation, so $g \circ f$ is well-defined. In fact, there's a formula:

Proposition 2.8. *If $f \leftrightarrow (a_n)$ and $g \leftrightarrow (b_n)$ with $a_0 = 0$, then $g \circ f \leftrightarrow (c_n)$, where*

$$c_n = \sum_{k \geq 0} b_k \sum_{\substack{n_1,\ldots,n_k \geq 1 \\ n_1+\cdots+n_k=n}} a_{n_1}\cdots a_{n_k}.$$

Proposition 2.9. *If $f \xleftrightarrow{\text{exp}} (a_n)$ and $g \xleftrightarrow{\text{exp}} (b_n)$ with $a_0 = 0$, then $g \circ f \xleftrightarrow{\text{exp}} (c_n)$, where*

$$c_n = \sum_{k \geq 0} \frac{b_k}{k!} \sum_{\substack{n_1,\ldots,n_k \geq 1 \\ n_1+\cdots+n_k=n}} \left(\frac{n!}{n_1!\cdots n_k!}\right) a_{n_1}\cdots a_{n_k}.$$

There is a good combinatorial interpretation of the coefficients of the composition of exponential generating functions. Imagine that you have two different kinds of "structures" that you can place on a finite set, and the number of these structures depends only on the size of the set. We'll call these mini-structures and meta-structures. If a_n is the number of mini-structures on any n-element set and b_n is the number of meta-structures on any n-element set, then c_n is the number of ways to take an n-element set, break it into nonempty pieces, place a mini-structure on each piece, and place a meta-structure on the set of pieces.

To see why this is, let X be an n-element set and $\Pi_k(X)$ denote the collection of all unordered set partitions of X into k nonempty subsets. The number of "mega-structures" described in the previous paragraph is

$$\sum_{k \geq 0} \sum_{\{B_1,\ldots,B_k\} \in \Pi_k(X)} b_k a_{|B_1|} a_{|B_2|} \cdots a_{|B_k|}. \tag{2.1}$$

Let $\overline{\Pi}_k(X)$ denote the set of ordered set partitions of X into k subsets. There are $k!$ ordered set partitions of X corresponding to each unordered set partition, so this sum is equal to

$$\sum_{k \geq 0} \frac{b_k}{k!} \sum_{(B_1,\ldots,B_k) \in \overline{\Pi}_k(X)} a_{|B_1|} a_{|B_2|} \cdots a_{|B_k|},$$

There are

$$\frac{n!}{n_1!\cdots n_k!}$$

ordered set partitions $(B_1,\ldots,B_k)$ of X which satisfy the equations $|B_i| = n_i$, so we can restate (2.1) as

$$\sum_{k \geq 0} \frac{b_k}{k!} \sum_{\substack{n_1,\ldots,n_k \geq 1 \\ n_1+\cdots+n_k=n}} a_{n_1} a_{n_2} \cdots a_{n_k},$$

which is of course the same as Proposition 2.9.

Example 2.10. The number of labelled trees with n vertices (here labelling means assigning each element of $[n]$ to exactly one vertex) is n^{n-2}. This is Cayley's formula; we'll talk about it later. Let $f \xleftrightarrow{\text{exp}} (n^{n-2})_{n \geq 1}$. Because $e^x \xleftrightarrow{\text{exp}} (1)_{n \geq 0}$, the coefficient of x^n in $e^{f(x)}$ is the number of ways to take the set $[n]$, break it up into pieces, draw a labelled tree in each piece, and then place the unique "empty meta-structure" over it all. But this is simply a labelled forest, so $e^{f(x)}$ is the generating function for the number of labelled forests with n vertices.

Example 2.11. The coefficients of the exponential generating function $e^{e^x - 1}$ are the Bell numbers, since it counts the number of ways to split $[n]$ into several sets and then do nothing with them.

From these two examples, we can see that the case $g = e^x$ is particularly important. This is the function we'll need to use if we want to count the number of total structures from the number of connected ones, for example.

Theorem 2.12 (Exponential formula). *If $f \xleftrightarrow{\text{exp}} (a_n)$ counts the number of connected structures that can be placed on an n-element set, then $[x^n]e^{f(x)}$ is the total number ways to divide an n-element set into connected structures.*

Exercise 2.2. Convince yourself of a similar interpretation of composition for ordinary generating functions: Show that if $f \leftrightarrow (a_n)$ counts the mini-structures and $g \leftrightarrow (b_n)$ counts the meta-structures, then $g \circ f \leftrightarrow (c_n)$ counts the number of ways to divide $[n]$ into subsets of consecutive integers, place a mini-structure on each subset, and place a meta-structure on the collection of subsets.

2.2. lagrange inversion

There is a theorem in complex analysis:

Theorem 2.13 (Complex Lagrange inversion). *If $f = \sum_{n=1}^{\infty} a_n x^n$ is analytic at $x = 0$ with $a_1 \neq 0$, then it has a unique compositional inverse function $g = \sum_{n=1}^{\infty} b_n x^n$ that is analytic at $x = 0$, and*

$$b_n = \frac{1}{n}[x^{n-1}]\left(\frac{x}{f(x)}\right)^n.$$

It turns out the same is true if you remove the condition of analyticity, and it's easier to prove, because you don't need any complex analysis. To do that, we'll reformulate the theorem.

Theorem 2.14 (Combinatorial Lagrange inversion). *If $R(t) = \sum_{n=0}^{\infty} r_n t^n$ and $r_0 \neq 0$, then the equation $f(x) = x \cdot R\big(f(x)\big)$ has a unique solution whose coefficients are*

$$[x^n]f = \frac{1}{n}[t^{n-1}]R(t)^n.$$

Proof of equivalence. Assume that complex Lagrange inversion (without analyticity) holds. Choose an $R(t)$ with $r_0 \neq 0$ and set $f = \frac{t}{R(t)}$. An inverse to f is a function g so that $f \circ g = \frac{g(t)}{R(t)} = t$. (Since $r_0 \neq 0$, there is a multiplicative inverse for $R(t)$.) But this means that $g(t) = tR\big(g(t)\big)$ and there is a solution to this. That g is the compositional inverse to f and you can check that the coefficients are correct.

Now suppose that the combinatorial Lagrange inversion holds and take any $f \leftrightarrow (a_n)$ with $a_0 = 0$ and $a_1 \neq 0$. We set $R(t) = \frac{1}{f(t)/t}$ and receive a function g such that $g(t) = tR\big(g(t)\big) = t\frac{g(t)}{f(g(t))}$, so $\frac{f(g(t))}{g(t)}g(t) = t$; or, more simply, $f(g(t)) = t$. So g is the compositional inverse of f. □

Why don't we also need to show that $g(f(t)) = t$? Well, it's not too hard to show, using Proposition 2.8, that any generating function f with $[x^0]f = 0$ and $[x^1]f \neq 0$ has a left inverse and also a right inverse. Then, necessarily, these two are equal.[2] So f always has a full inverse, and it's unique.

All told, though, Theorem 2.14 seems to be a pretty odd theorem, and it's not clear that it's actually very powerful. Before we prove it, it's worth seeing it in action.

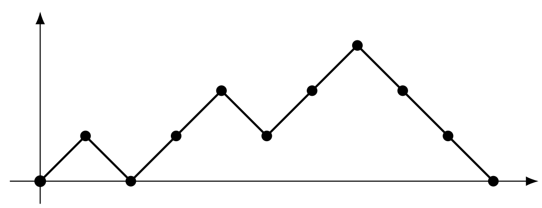
2.3. applications of lagrange inversion

Catalan numbers

Welcome to the wonderful world of Catalan numbers. What are they? Just like any other combinatorial sequence, they're the answer to a counting problem.

Definition 2.15. A Dyck path of length $2n$ is a sequence of n up steps and n down steps such that, reading left to right, there are never more down steps than up steps. The Catalan number C_n is the total number of Dyck paths of length $2n$.

Dyck paths are usually drawn on the plane with diagonal up and down steps. The sequence (up, down, up, up, down, up, up, down, down, down), for example, is drawn like this:



They have a *ton* of combinatorial interpretations. And by this I mean a *metric ton*. Maybe two. In *Enumerative Combinatorics*, Richard Stanley lists sixty-six different interpretations as exercises, but that wasn't nearly enough for him, so he made an appendix with even more. Let's just list a few.

Exercise 2.3. A *proper pairing* of $2n$ parentheses is a way to list n opening and n closing parentheses that could occur as an actual grouping; in other words, as you read left to right, there are never more closing parentheses than opening ones. Show that the number of proper pairings of $2n$ parentheses is C_n.

Definition 2.16. A *binary tree* is a tree with a specified *root vertex*; every vertex has at most two children, labelled "left" or "right." A *complete binary tree* is a binary tree where each vertex has either two children or none. (Vertices with no children are called leaves.)

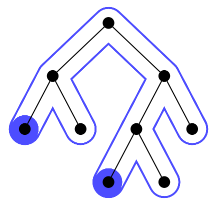
Example 2.17. Binary trees are usually drawn with the root at the top and children coming downward, sort of like genealogical trees. Binary trees look like this:

The right tree is complete, while the left tree is not.

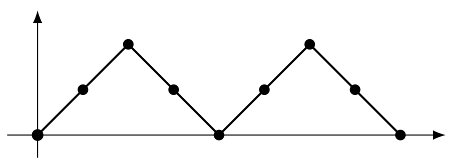
Proposition 2.18. *C_n is the number of complete binary trees with $n+1$ leaves.*

[2] This is just the usual argument: If g_1 is a left inverse and g_2 is a right inverse, then $g_1 = g_1 \circ f \circ g_2 = g_2$.

Proof. Draw a path around the complete binary tree starting at the root and going counterclockwise; for example:



Any complete binary tree with $n+1$ leaves has $2n$ edges; record the order of the left and right edges in this walk, ignoring repeats. For the example tree, the order is l, l, r, r, l, l, r, r: The first three edges in the walk are left, left; then the second edge is repeated and the next edge is right; then repeated edges and the next edge is right—and so on. This easily turns into a Dyck path by converting left edges to up steps and right edges to down steps. Here, we get



Any Dyck path can be converted to the corresponding binary tree by reversing the process, so this map is a bijection. □

The process that orders the edges in this proof is called *depth-first search*.

Exercise 2.4. Show that the number of binary trees with n vertices is C_n. (Hint: Construct a bijection to the set of complete binary trees with $n+1$ leaves.)

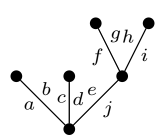
Definition 2.19. A *plane tree* is a rooted tree in which the children of each node are ordered.

Example 2.20. Here are two pictures that represent the same rooted tree, but not the same plane tree.

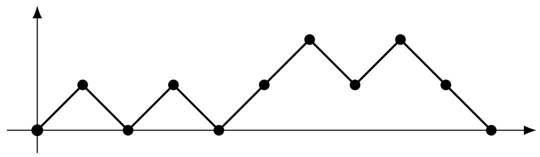
They are called *plane trees* because each corresponds to a drawing of the rooted tree in the plane, where the ordering is left-to-right.

Proposition 2.21. *The number of plane trees with $n+1$ vertices is C_n.*

Proof. Flip the tree upside down and perform a depth-first search, recording whether the edge is traversed downward or upward. For the tree



depth first search traces through the edges in the order $a, b, c, \ldots, i, j$. This is easily converted to a Dyck path: Convert each up step in the tree to an up step in the Dyck path; same for down steps. This tree gives the Dyck path



I leave it to you to prove that this map is in fact a bijection. $\square$

Anyway, we want to find a formula for the Catalan numbers, and we're going to do this using generating functions. The first ingredient is a recursive formula for the Catalan numbers.

Proposition 2.22. *The Catalan numbers are related by the recurrence*

$$C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}.$$

Proof sketch. Take a complete binary tree with $n+1$ leaves and delete the top vertex. You're left with two complete binary trees, one with k leaves and the other with $n+1-k$ leaves. This operation provides a bijection between the set of complete binary trees with $n+1$ leaves and the set of pairs of complete binary trees (T_1, T_2) where the total number of leaves amongst T_1 and T_2 is $n+1$. The number of such pairs where T_1 has k leaves is the right side of the equation. $\square$

Now: Let $C(x) \leftrightarrow (C_n)$. Take both sides of this recurrence and multiply by x^n to get

$$C_n x^n = x \sum_{k=1}^{n} \big(C_{k-1} x^{k-1}\big)\big(C_{n-k} x^{n-k}\big);$$

then we sum over all $n \geq 1$:[3]

$$\begin{aligned}
\sum_{n=1}^{\infty} C_n x^n &= x \sum_{n=1}^{\infty} \sum_{k=1}^{n} \big(C_{k-1} x^{k-1}\big)\big(C_{n-k} x^{n-k}\big) \\
C(x) - 1 &= x \sum_{\substack{k-1 \geq 0 \\ n-k \geq 0}} \big(C_{k-1} x^{k-1}\big)\big(C_{n-k} x^{n-k}\big) \\
C(x) - 1 &= x \cdot C(x) \cdot C(x).
\end{aligned}$$

At this point, we could write that $xC(x)^2 - C(x) + 1 = 0$ and use the quadratic formula to conclude that

$$C(x) = \frac{1 - \sqrt{1-4x}}{2}.$$

(We have to choose the minus sign so that the expression has a power series at $x = 0$.) Then we could expand this into a power series and determine the coefficient of x^n. But this feels a bit sketchy (though it can be made rigorous via some analytic power series–generating function correspondence) and the calculation is pesky. Using Lagrange inversion is easier.

The expression $C(x) - 1 = xC(x)^2$ is very similar to the form $C(x) - 1 = xR\big(C(x)\big)$ except for that pesky 1. So we get rid of it by defining $\tilde{C}(x) = C(x) - 1$; then

$$\tilde{C}(x) = x\big(\tilde{C}(x) + 1\big)^2.$$

[3] We can't include $n = 0$, because there's no recurrence when $n = 0$.

So take $R(t) = (1+t)^2$. By Theorem 2.14, we have, for $n \geq 1$,

$$C_n = [x^n]\tilde{C}(x) = \frac{1}{n}[t^{n-1}]R(t)^n = \frac{1}{n}[t^{n-1}](1+t)^{2n} = \frac{1}{n}\binom{2n}{n-1}.$$

This is the same as $\frac{1}{n+1}\binom{2n}{n}$, which is how the formula is usually expressed. And this expression is true for $n = 0$, too. So:

Theorem 2.23. $C_n = \frac{1}{n+1}\binom{2n}{n}$.

There's another benefit to Lagrange inversion beyond aesthetics: It generalizes very easily. Let C_n^m denote the number of complete m-ary trees with $n+1$ nodes; you can check that these numbers satisfy a similar recurrence relation as C_n, and their generating function $C^m(x)$ satisfies

$$C^m(x) - 1 = x^{m-1}\big(C^m(x)\big)^m \tag{2.2}$$

Since any complete m-ary tree has $n \equiv 1 (\text{mod } m-1)$ leaves, the number $C_n^m \neq 0$ if and only $m - 1 \mid n$. So the $\tilde{C}^m(x) := C(x^{1/(m-1)}) - 1$ is a power series that satisfies the equation $\tilde{C}^m(x) = x\big(\tilde{C}^m(x) + 1\big)^m$. Setting $R(t) = (1+t)^m$ solves the problem again, with

$$[x^n]\tilde{C}(x) = \frac{1}{n}[t^{n-1}](1+t)^{mn} = \frac{1}{n}\binom{mn}{n-1},$$

so

$$C_n^m = \begin{cases} 0 & \text{if } m-1 \nmid n \\ \frac{1}{k}\binom{mk}{k-1} & \text{if } n = k(m-1). \end{cases}$$

In contrast, it seems much more difficult to try to solve (2.2) algebraically, and even if you could, an absolute pain to expand out analytically and then extract the coefficients. Lagrange inversion is much simpler.

A different proof of this formula for C_n uses the fact that C_n is the number of plane trees with $n+1$ vertices.

Exercise 2.5. Let $f \leftrightarrow (a_n)$, where a_n is the number of plane trees with n vertices.

1. Prove that $f(x) = x\frac{1}{1-f(x)}$. (Here, $\frac{1}{1-x}$ means the multiplicative inverse of $1-x$, which is the power series $1 + x + x^2 + \cdots$.)
2. Use Lagrange inversion to show that $C_n = [x^{n+1}]f = \frac{1}{n+1}\binom{2n}{n}$.

Rooted labelled trees and Cayley's formula

Definition 2.24. A *rooted labelled tree* is, naturally enough, a labelled tree in which one vertex is declared the root. A *planted forest* is a labelled forest in which each tree has a root. We'll denote the number of rooted labelled trees with n vertices by $\mathrm{rt}(n)$ and the number of planted forests with n vertices by $\mathrm{pf}(n)$.

If we let $f \overset{\text{exp}}{\longleftrightarrow} \big(\mathrm{rt}(n)\big)$ and $g \overset{\text{exp}}{\longleftrightarrow} \big(\mathrm{pf}(n)\big)$, then the Exponential Formula tells us that

$$g(x) = e^{f(x)}.$$

Moreover, the coefficient of $x^n/n!$ in $xe^{f(x)}$ is $n\ \mathrm{pf}(n-1)$; it is the number of ways to choose a single vertex and then place a planted forest on the remaining nodes. But this exactly corresponds to a rooted labelled tree: Declare the single vertex the root, and then connect it to the roots of the components in the forest. This is a bijection (the inverse simply consists of taking a rooted labelled tree, deleting the edges incident to the node, and declaring as the new roots of the forest each vertex that had been adjacent to the old root). This means that

$$f(x) = xe^{f(x)}.$$

The coefficients of f are therefore easily extractable using Lagrange inversion with $R(t) = e^t$:

$$\frac{\mathrm{rt}(n)}{n!} = \frac{1}{n}[t^{n-1}]e^{nt} = \frac{1}{n}\frac{n^{n-1}}{(n-1)!} = \frac{n^{n-1}}{n!}.$$

So the number of rooted trees with n vertices is $\mathrm{rt}(n) = n^{n-1}$. Of course, the number of rooted labelled trees is exactly n times the number of labelled trees, so we've just proved Cayley's formula:

Theorem 2.25. *The number of distinct labelled trees with n vertices is n^{n-2}.*

And it only took half a page.

2.4. proof of lagrange inversion

Let $f \leftrightarrow (a_n)$. There are two conditions:

$$a_n = [x^n]xR\big(f(x)\big) = [x^n - 1]R\big(f(x)\big) \tag{$*$}$$

$$a_n = \frac{1}{n}[t^{n-1}]R(t)^n \tag{$**$}$$

We want to show that condition ($*$) and condition ($**$) are equivalent. To do this, we transform each condition into a much more formidable-looking sum and then show that these are the same. Let's jump in.

We'll start with condition ($*$). It says that

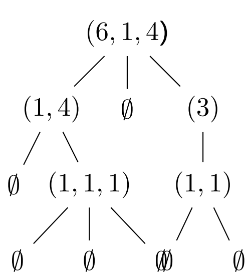
$$a_n = [x^{n-1}]\big(r_0 + r_1 f(x) + r_2 f(x)^2 + \cdots\big),$$

or

$$a_1 = r_0 \qquad \text{and} \qquad a_n = \sum_{k=0}^{\infty} \sum_{\substack{n_1,\ldots,n_k \geq 1 \\ n_1+\cdots+n_k=n-1}} r_k a_{n_1} \cdots a_{n_k}.$$

Now, each of the a_{n_i} is itself equal to a recurrence, so we can imagine plugging that in, which gives us a formula in terms of a_i with smaller indices, and then using it again and again until we're left with a formula for a_n that uses only the r_k.

Now, that's a lot to keep track of, so it helps to think about it in steps. In the formula above, each term corresponds to a composition of $n-1$. When we substitute the recursion for the first time and then expand out, each term corresponds to a pair: a composition $(n_1, \ldots, n_k)$ of $n-1$, and for each i, a composition of $n_i - 1$. And so on. It helps to keep track of these steps in a diagram. Here's an example for $n = 12$:



But wait just a minute, you say. That's a tree!

Good eye. Every term in the "unravelling" of the recurrence corresponds to a unique "composition tree" like this one. If we remove the labels from this tree, we get a plane tree. In fact, each plane tree corresponds to a *unique* composition tree. You'll notice that in the tree above, there are 6 vertices in the left branch, 1 in the middle, and 4 in the right. This is not a coincidence: In any vertex of a composition tree, the label of that node tells you the number of vertices in each branch that flows from it. So we can tell exactly what the compositions on the vertices of a plane tree must be just from its structure.

Now, what's the value of the term that the plane tree represents? It's just the product of a bunch of r_k's, where k is the number of terms in the composition. The tree shown above has two compositions into three parts, two into two parts, one into one part, and seven into no parts, so it corresponds to the term $r_3^2 r_2^2 r_1^1 r_0^7$. Given a plane tree T, we define the expression

$$r^T = \prod_{v \in V(T)} r_{\mathrm{ch}(v)},$$

where $\mathrm{ch}(v)$ is the number of children of v. What this all means is that we can write condition ($*$) in the very bizarre form

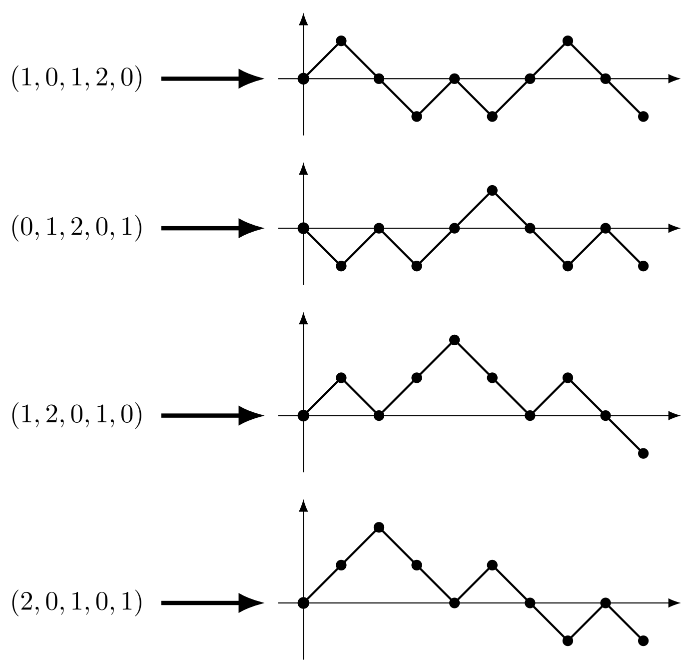
$$a_n = \sum_{\substack{\text{plane trees}\\ T \text{ on } n\\ \text{vertices}}} r^T. \tag{$\star$}$$

Now that we've "simplified" the first condition, let's turn to ($**$). Under this assumption, the coefficients are given by

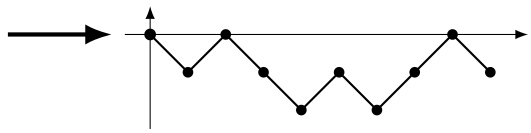
$$a_n = \frac{1}{n} \sum_{\substack{c_1,\ldots,c_n \geq 0\\ c_1+\cdots+c_n=n-1}} r_{c_1} \cdots r_{c_n}.$$

The terms in this sum are indexed by weak compositions of $n-1$ into n parts. Consider the equivalence classes of these compositions generated by cyclic shifts. Each equivalence class must contain exactly n compositions: If the equivalence class of $(c_1, \ldots, c_n)$ has k compositions, then $k \mid n$ and $\frac{n}{k} \mid c_1 + \cdots c_n = n-1$. The only way for this to be true is if $k = n$.

So after distributing the $\frac{1}{n}$ factor, the terms are indexed by cyclic equivalence classes of compositions of $n-1$. We can turn each composition $(c_1, \ldots, c_n)$ into a Dyck-type path composed of c_1 up steps followed by one down step, c_2 steps followed by one down step, ... For example,



$(0, 1, 0, 1, 2)$



All of these paths necessarily end at the height $y = -1$. But among them, there is exactly one that remains above the x-axis at all previous steps. It turns out that this is true for each equivalence class.

Exercise 2.6. Let $(c_1, \ldots, c_n)$ be a composition of $n-1$. Suppose that, among the lowest points in the Dyck-type drawing of this composition, the leftmost one follows the ith down step. Show that the Dyck-type drawing of $(c_{i+1}, \ldots, c_n, c_1, \ldots, c_i)$ is a Dyck path with an added down step at the end. Then show that this is the only cyclic permutation of the composition with this property.

Therefore each term in the sum corresponds to a Dyck path with $2n-2$ steps with an added final down step. And it's clear we can go backwards: Each such path corresponds to a term in the sum. Given a path P of this form, let P_{up} be the n-element vector $(p_1, \ldots, p_n)$ where p_i is the number of up steps that immediately precede ith down step in P. Let's define

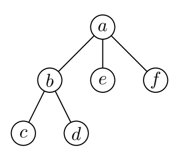
$$r^P = \prod_{i=1}^{n} a_{p_i}.$$

What we have, then, is that condition ($**$) is equivalent to the statement that

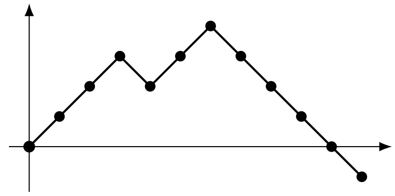
$$a_n = \sum_{\substack{\text{Dyck paths } P \\ \text{with } 2n \quad 2 \text{ steps} \\ \text{plus final down step}}} r^P. \tag{$\star\star$}$$

To finish the proof, we want to show that ($\star$) and ($\star\star$) are the same. To do this, we provide a bijection between plane trees and modified Dyck paths $T \mapsto P$ such that $r^T = r^P$. If we find such a bijection, then the terms of ($\star$) and ($\star\star$) correspond exactly.

We gave a bijection before between plane trees and Dyck paths, but that bijection doesn't preserve the monomial r^T. If we modify it a bit, we can make it work. This time, though, instead of paying attention to the edges, pay attention to the order in which you first encounter the vertices. Depth-first search on the plane tree



encounters the vertices in the order $(a, b, c, b, d, b, a, e, a, f, a)$. Removing repeats, we get an ordered list of first encounters: (a, b, c, d, e, f). We build a modified Dyck path from the list $(v_1, \ldots, v_k)$ by drawing $\mathrm{ch}(v_1)$ up steps followed by a down step, then $\mathrm{ch}(v_2)$ up steps followed by a down step, and so on. You can check that, starting from a plane tree with n vertices, this will have exactly $n-1$ up steps and n down steps (because each vertex contributes one down step, and every vertex but the root contributes one up step). The modified Dyck path coming from the tree above is



Exercise 2.7. Convince yourself that this map always produces a modified Dyck path. (Explain why it must stay above the x-axis until the last step.) Then show that any modified Dyck path corresponds to exactly one plane tree under this map.

Moreover, this bijection preserves the polynomials r^T and r^P. Suppose that $T \mapsto P$ under this bijection. If depth-first search on T encounters its vertices in the order $(v_1, \ldots, v_n)$ and $P_{\mathrm{up}} = (p_1, \ldots, p_n)$, then $\mathrm{ch}(v_i) = p_i$. Therefore

$$r^T = \prod_{i=1}^{n} a_{\mathrm{ch}(v_i)} = \prod_{i=1}^{n} a_{p_i} = r^P.$$

So ($\star$) and ($\star\star$) are equivalent, which means that ($*$) and ($**$) are, as well. □

2.5. an extension of lagrange inversion

Lagrange inversion can be extended to the following theorem:

Theorem 2.26. *If $R(t) = \sum_{n=0}^{\infty} r_n t^n$ and $r_0 \neq 0$, then the equation $f(x) = x \cdot R\big(f(x)\big)$ has a unique solution, and*

$$[x^n] f(x)^k = \frac{k}{n} [t^{n-k}] R(t)^n.$$

The proof requires only a slight modification of the argument from above.

Proof sketch. Let $f \leftrightarrow (a_n)$. We have

$$[x^n] f(x)^k = \sum_{\substack{n_1, \ldots, n_k \geq 0 \\ n_1 + \cdots + n_k = n}} a_{n_1} \cdots a_{n_k}. \tag{2.3}$$

When we unfolded the recurrence in the previous proof, each term was represented by a plane tree. Each term in (2.3) is therefore represented by a *plane forest* of k trees: An ordered set $F = (T_1, \ldots, T_k)$ where each T_i is a plane tree.

On the other side, consider

$$[x^{n-k}] R(x)^n = \sum_{\substack{c_1, \ldots, c_n \geq 0 \\ c_1 + \cdots + c_n = n-k}} r_{c_1} \cdots r_{c_n}.$$

So terms in this sum are indexed by compositions of $n-k$ into n parts, each of which corresponds to a Dyck-type path with $n-k$ up steps and n down steps, ending on the line $y = -k$.

The map from plane trees to Dyck-type paths ending on $y = -1$ can be extended to plane forests by simply enumerating the vertices in a forest first by the overall order of the trees and then by the depth-first order within. (So if the vertices in T_1 are ordered as $(v_1, \ldots, v_n)$ by depth-first search and the vertices of T_2 by $(u_1, \ldots, u_m)$, then the vertices of $F = (T_1, T_2)$ are ordered as $(v_1, \ldots, v_n, u_1, \ldots, u_m)$.) Every Dyck-type path in the image of this map has $n-k$ up steps and n down steps and the additional property that the only vertex on the line $y = -k$ is the final one. The converse is also true: Every such path corresponds to exactly one plane forest.[4] For the purposes of this proof, we will call such paths *proper*.

Now fix a composition $(c_1, \ldots, c_n)$ of $n-k$ and consider the Dyck-type paths corresponding to the elements of its equivalence class under cyclic permutation. Exactly k of these paths are proper:

[4] Take a path P and mark the leftmost vertex on each horizontal line below the x-axis. Each tree in the forest that maps to P corresponds to a segment between two marked vertices.

Exercise 2.8. Suppose that $(c_1, \ldots, c_n)$ is a composition of $n-k$, and suppose the leftmost point in its Dyck-type path on the line $y = -j$ follows the i_jth down step. Show that in the Dyck-type path of $(c_{i_j+1}, \ldots, c_n, c_1, \ldots, c_{i_j})$, the only vertex on the line $y = -k$ is the final one. Then show that every other permutation of the composition violates this property.

In particular: If $F = (T_1, \ldots, T_k)$ corresponds $C = (c_1, \ldots, c_n)$, then the cyclic shifts of F correspond to the cyclic shift of C whose associated Dyck-type path is proper. If there are n/m distinct cyclic shifts of C, then there are k/m distinct cyclic shifts of F. So there are exactly n/k times as many compositions as there are plane forests.

Recalling that $r^F = r^C$, we have

$$[x^n]f(x)^k = \sum_{\substack{F=(T_1,\ldots,T_k)\\ |T_1|+\cdots+|T_k|=n}} r^F = \frac{k}{n} \sum_{\substack{C=(c_1,\ldots,c_n)\\ c_i \geq 0 \\ c_1+\cdots+c_n=n-k}} r^C = \frac{k}{n}[x^{n-k}]R(x)^n. \qquad \square$$

2.6. statistics

Definition 2.27. A *combinatorial statistic* on a set $\mathcal{S}$ is a function $\varphi\colon \mathcal{S} \to \mathbb{N}_0$.

We think of $\mathcal{S}$ as being a class of combinatorial objects and φ as being a measure of something or, well, statistic on the elements of that class. For example, the function $G \mapsto |E(G)|$ that sends a graph to its number of edges is a simple statistic. So are the maps to its number of vertices, maximum degree, diameter, girth, and so on.

If $\varphi^{-1}(k)$ is finite for every $k \in \mathbb{N}_0$, then we can define its generating function

$$F_\varphi = \sum_{a \in \mathcal{S}} q^{\varphi(a)}.$$

If we take $\mathcal{S}$ to be the set of connected graphs, for example, then the statistic $G \mapsto |E(G)|$ has a generating function.

Definition 2.28. Two statistics $(\mathcal{S}, \varphi)$ and $(\mathcal{T}, \psi)$ are *equidistributed* if $F_\varphi = F_\psi$.

For example:

Proposition 2.29. *The following statistics on the collection of plane trees with $n+1$ vertices and the collection of Dyck paths with $2n$ steps are equidistributed.*

1. *The degree of the root vertex of the plane tree.*
2. *The number of initial consecutive up steps in the Dyck path.*
3. *The level of the leftmost vertex in a plane tree.*
4. *One less than the number of times a Dyck path hits the x-axis.*

Proof. That (1) and (2) are equidistributed follows from the bijection between plane trees and Dyck paths presented in Section 2.4; the bijection in Section 2.3 shows that (2) and (3) are equidistributed, and it also shows that (1) and (4) are equidistributed. $\square$

Definition 2.30. The *ballot number* $B(n,k)$ is the number of Dyck paths with $2n$ steps and an initial run of k up steps. (And therefore also any of the equidistributed statistics in Proposition 2.29.)

Proposition 2.31. *The following statistics on the collection of plane trees with $n+1$ vertices and the collection of Dyck paths with $2n$ steps are equidistributed.*

1. *The number of peaks in a Dyck path.*
2. *The number of leaves in a plane tree.*
3. *The number of non-leaves in a plane tree.*

4. *One more than the number of double descents (two consecutive downsteps) in a Dyck path.*

Proof. The bijection from Section 2.3 shows that (1) and (2), as well as (3) and (4), are equidistributed; the bijection from Section 2.4 shows that (2) and (4) are. □

Definition 2.32. The *Narayana numbers* $N(n,k)$ are the number of Dyck paths with $2n$ steps and k peaks. (Or any of the three other interpretations in Proposition 2.31.)

This is enumerative combinatorics: Let's find the formulas!

Proposition 2.33. $B(n,k) = \frac{k}{n}\binom{2n-k-1}{n-k}$.

Proof. If $f(x)$ is the generating function for the number of plane trees with n vertices, then $f(x)^k$ is the generating function for the number of plane forests with n vertices and k trees—which is equal to the number of plane trees with $n+1$ vertices. Recall that $f(x) = x\frac{1}{1-f(x)}$ (see Exercise 2.5), so our supercharged Lagrange inversion gives

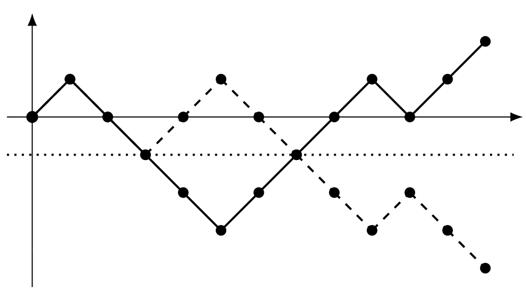
$$B(n,k) = [x^n]f(x)^k = \frac{k}{n}[x^{n-k}]\left(\frac{1}{1-x}\right)^n = \frac{k}{n}\binom{2n-k-1}{n-k},$$

since $(\frac{1}{1-x})^n$ is the generating function for the number of compositions into n parts. □

If you prefer, here's a more direct proof of the formula.

Second proof of Proposition 2.33. If we horizontally reflect a Dyck path that has an initial run of k up steps, we get a Dyck path that ends with one up step followed by k down steps. Deleting this segment results in a Dyck path from the origin to $(2n-k-1, k-1)$ that remains above the x-axis. This operation is a bijection, so we'll instead count the number of such paths.

If we remove the condition that the path remains above the x-axis, then there are $\binom{2n-k-1}{n-1}$ different paths (we just need to specify where the $n-k$ down and $n-1$ up steps are). Say a path is *bad* if it intersects the line $y=-1$. For every such path, we can find the first vertex that intersects with the line $y=-1$ and reflect all of the path to the right across this line, like this:



This is a bijection between the set of Dyck paths from the origin to $(2n-k-1, k-1)$ and $(2n-k-1, -k-1)$. Since there are $\binom{2n-k-1}{n}$ such paths, we have

$$B(n,k) = \binom{2n-k-1}{n-1} - \binom{2n-k-1}{n} = \binom{2n-k-1}{n-1} - \frac{n-k}{n}\binom{2n-k-1}{n-1} = \frac{k}{n}\binom{2n-k-1}{n-1},$$

which is equal to $\frac{k}{n}\binom{2n-k-1}{n-k}$. □

Exercise 2.9. Use this technique to provide a direct proof of the formula $C_n = \frac{1}{n+1}\binom{2n}{n}$ by counting the number of Dyck paths with $2n$ steps.

Proposition 2.34. $N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}$.

Proof. We know that if $f \leftrightarrow (a_n)$ and $f(x) = xR\big(f(x)\big)$, then

$$a_{n+1} = \sum_{\substack{\text{Dyck paths } P \\ \text{with } 2n \text{ steps} \\ \text{plus final down step}}} r^P.$$

We'll write P^0 to denote the Dyck path obtained from P by removing the final down step. The number of times r_0 appears in the expression r^P is one more than the number of double descents in P^0. So if we set $R(t) = q + t + t^2 + \cdots$, then the solution to $f(x) = xR\big(f(x)\big)$ is a power series with the coefficients

$$a_{n+1} = \sum_{k=1}^{n} N(n,k)q^k.$$

Before you cry foul—an indeterminate q isn't a number!—look back at our proof of Lagrange inversion. All we really used is that they're members of a commutative ring, so certainly the ring $\mathbb{Z}[q]$ works. Anyway: For this generating function,

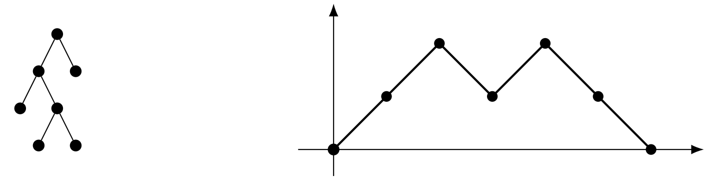
$$\begin{aligned}
a_{n+1} &= \frac{1}{n+1}\ [t^n] \left(q + \frac{t}{1-t}\right)^{n+1} \\
N(n,k) &= \frac{1}{n+1}\ [q^k][t^n] \left(q + \frac{t}{1-t}\right)^{n+1} \\
&= \frac{1}{n+1}\binom{n+1}{k}\ [t^n] \left(\frac{t}{1-t}\right)^{n+1-k} \\
&= \frac{1}{n+1}\binom{n+1}{k}\ [t^{k-1}] \left(\frac{1}{1-t}\right)^{n+1-k} \\
&= \frac{1}{n+1}\binom{n+1}{k}\binom{n-1}{k-1}.
\end{aligned}$$

And some binomial calculus turns this into $\frac{1}{n}\binom{n}{k}\binom{n}{k-1}$. □

For fun, here's one more interpretation of the Narayana numbers.

Proposition 2.35. *There are exactly $N(n, k-1)$ binary trees with n vertices and k right edges.*

Proof. We first present a bijection between the complete binary trees with $n+1$ leaves and the Dyck paths with $2n$ steps. Any complete binary tree with $n+1$ leaves has exactly $2n$ edges. Given a complete binary tree T, we can perform a depth-first search and record the order in which the edges are encountered. Then, this list can be transformed into a Dyck path by translating left and right edges to up and down steps, respectively. For example, the tree on the left is taken to the Dyck path on the right:



The left edge of any vertex precedes the right edge in the depth first search, so the path that this mapping produces never crosses below the x-axis. Conversely, it's possible to write down a complete binary tree that maps to it; so this correspondence is a bijection which we denote by f.

We also have a bijection between complete binary trees with $n+1$ vertices and binary trees with n vertices by deleting all the leaves. Denote the inverse of this bijection by g. We will focus on the bijection $h = f \circ g$ that takes in a binary tree with n vertices and returns a Dyck path with $2n$ steps. We will show that the number of peaks of $f(T)$ is exactly one more than the number of right edges in T, which proves the desired statement.

Let B be a complete binary tree. When $g(B)$ is constructed from B, a right edge in B corresponds to the last down step before an up step if and only if that edge is internal (i.e., not incident to a leaf). This is because (1) the child vertex of any internal edge has a left edge, which is next in the depth-first search; and (2) any right edge incident to a leaf is followed in the depth-first search by another right edge. This means that the set of right edges in a binary tree T are in exact correspondence to the set of down steps in $h(T)$ that immediately precede an up step. This number is one less than the number of peaks in $h(T)$. So if T has $k-1$ right edges, then $h(T)$ has k peaks. □

3. a handful of sequences

3.1. labelled trees

We've looked at the number of labelled trees; now we look at labelled binary and plane trees.

Definition 3.1. A *labelled binary tree* is a binary tree on the vertex set $[n]$. (In other words, each vertex has a label between 1 and n, the number of vertices, and no label repeats.)

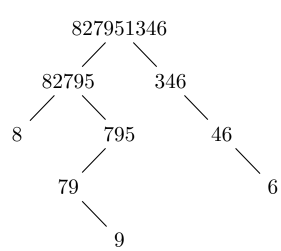
The number of binary trees with n vertices is C_n (see Exercise 2.4). Any way that we assign labels creates a different labelled binary tree, so the total number is just $n!\,C_n$. Things become more interesting when we start adding restrictions to the labellings.

Definition 3.2. A (labelled) binary tree is called *increasing* if the label of every child is less than the label of its parent.

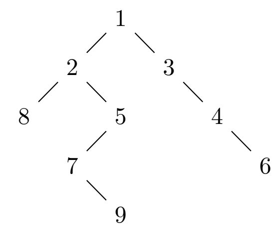
Proposition 3.3. *There are $(n-1)!$ increasing binary trees with n vertices.*

Proof sketch. Deleting the vertex labelled n is a surjective map from the set of n-vertex increasing binary trees to the set of $(n-1)$-vertex increasing binary trees. Moreover, a binary tree with $n-1$ vertices has n places to add a leaf, so the preimage of every element contains exactly n trees. Now induct, using the fact that there is $1 = 1!$ increasing binary tree with 1 vertex. □

Here's an explicit bijection from S_n to the set of increasing binary trees, explained through example. Take a permutation, say $827951346 \in S_9$ (this is the map $1 \mapsto 8$, $2 \mapsto 2$, $3 \mapsto 7$, etc.), and split it into two sequences at the 1. Then split each of these sequences at the lowest number they contain, and so on. We can draw this as a tree like this:

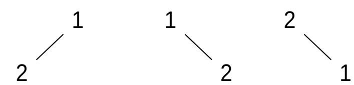


If, instead of writing the whole permutation, we just record the number at which we split, we get an increasing binary tree:



Exercise 3.1. Show that this is a bijection.

On to the next labelling: A labelled binary tree is called *left-increasing* if the label of every left child is greater than the label of its parent. There are three left-increasing binary trees with 2 vertices:



You can check that there are 16 left-increasing binary trees with 3 vertices.

Problem 3.2. Prove that there are $(n+1)^{n-1}$ left-increasing binary trees with n vertices.

All of the classes "left-decreasing," "right-increasing," and "right-decreasing" have the same number of trees as left-decreasing. (Why?) So we're left with one last class.

Problem 3.3. Show that the number of left-decreasing right-increasing trees with 1, 2, and 3 vertices are 1, 2, and 7, respectively. Then find a formula for these numbers.

Definition 3.4. A *binary search tree* is a labelled binary tree in which, for every vertex v, the label of every vertex in the left subtree of v is less than the label of v, and the label of every vertex in the right subtree of v is greater than the label of v.

Exercise 3.4. Show that the number of binary search trees with n vertices is C_n. (**Hint:** How many binary search trees are there on a given (unlabelled) binary tree?)

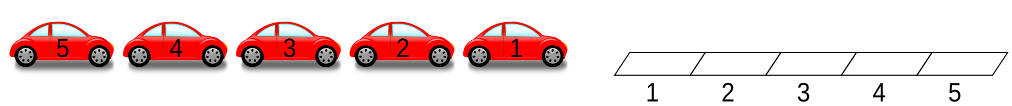
Finally, we turn to plane trees.

Problem 3.5. Show that the number of increasing labelled plane trees with 1, 2, and 3 vertices is 1, 3, and 15, respectively. Find a formula for these numbers.

3.2. parking functions

It's time to introduce a new combinatorial object. You know what that means: Story time.

Imagine that there are n cars, labelled 1 through n from front to back, in a queue for a string of n parking spots. The situation looks like this:[1]



Each car i has a favorite parking space a_i, which we can record in the sequence $(a_1, \ldots, a_n)$. Here's how these silly drivers park: First, they drive to their favorite parking space. If it's already taken, they can't turn back—it's a one-way road—so they keep driving and take the first open space. If they don't find any open parking spaces, then they keep driving off the end of the page and into oblivion. A *parking function* is a sequence in which every car finds a spot to park. (It's called a function because we can just as well think of this sequence as a function $[n] \to [n]$.)

[1] I couldn't find a good LaTeX car icon, so I just used one of the first clipart pictures I found—super classy.

Exercise 3.6. Show that any permutation $(\sigma_1, \dots, \sigma_n) \in S_n$ is a parking function.

If $(a_1, \dots, a_n)$ is a parking function, then there are at most k drivers whose favorite spot is among the last k. In other words, it's a necessary condition that, for every $k \in [n]$,

$$\Big|\{i \in [n] : a_i \geq n+1-k\}\Big| \leq k. \tag{$*$}$$

It turns out that this completely characterizes parking functions.

Exercise 3.7. Show that any sequence $(a_1, \dots, a_n)$ that satisfies condition ($*$) is a parking function.

There is also a different characterization using permutations:

Exercise 3.8. Show that $(a_1, \dots, a_n)$ is a parking function if and only if there is a permutation $\sigma \in S_n$ such that $a_i \leq \sigma_i$ for every $i \in [n]$.

And now the question: How many parking functions are there? It's helpful to think of the problem in a new way. Instead of arranging n parking spots in a line, let's arrange $n+1$ parking spots in a circle. The cars enter at parking spot 1, drive to their preferred spot, and continue around the circle until the first available spot. In this situation, every car is always able to park, and there's one spot left over.

A function $[n] \to [n]$ is a parking function in this new problem if and only if no car ends up in spot $n+1$ at the end. Now, notice: If we partition the functions $[n] \to [n+1]$ into equivalence classes by cyclic permutations, each equivalence class contains exactly one parking function. Every equivalence class contains $n+1$ functions, so:

Proposition 3.5. *There are $(n+1)^{n-1}$ parking functions on an n-element set.*

It's the Cayley numbers again! These numbers pop up in all sorts of places.

Exercise 3.9. We can extend the parking situation to a scenario with n cars and s spots to define *s-parking functions* $[n] \to [s]$. (We assume that $s \geq n$.) Show that there are $(s+1-n)(s+1)^{n-1}$ parking functions $[n] \to [s]$.

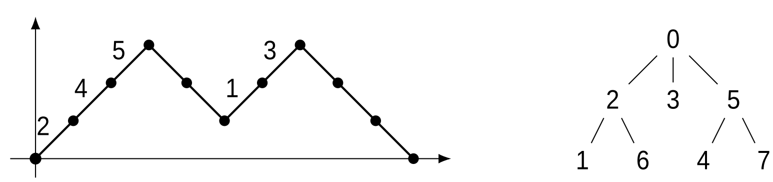
3.3. the cayley numbers $(n+1)^{n-1}$

Some further places in which the Cayley numbers appear:

Definition 3.6. A plane tree whose $n+1$ vertices are the set $\{0, 1, \dots, n\}$ is called *child-increasing* if the root is labelled with 0 and, for every vertex with children, the labels of its children increase from left to right.

Definition 3.7. A *labelled Dyck path* with $2n$ steps has each of its up steps labelled with $\{1, 2, \dots, n\}$, with no repeats, such that the labels in every consecutive set of up steps are increasing.

A child-increasing plane tree and a labelled Dyck path:



Let's sum up.

Proposition 3.8. *There are $(n+1)^{n-1}$ elements in each of the following collections:*

1. *labelled trees with* $n+1$ *vertices,*
2. *parking functions* $[n] \to [n]$*,*
3. *left-increasing binary trees with* n *vertices,*
4. *equivalence classes of cyclic shifts of functions* $[n] \to [n+1]$*,*
5. *child-increasing plane trees with* $n+1$ *vertices, and*
6. *labelled Dyck paths with* $2n$ *steps.*

Proof. (1) is Theorem 2.25, (2) is Proposition 3.5, (3) is Problem 3.2, and (4) is from the fact that the set has $(n+1)^n$ elements and each equivalence class has $n+1$. Given a labelled tree with $n+1$ vertices, there is exactly one way to draw it as a child-increasing plane tree, which shows that (1) and (5) have the same size. And the second bijection between plane trees and Dyck paths (used in Section 2.4) can be extended to labelled versions by simply writing the labels of the children of a vertex on the corresponding set of up steps; this shows that (5) and (6) have the same size. □

4. polytopes

4.1. the beginning

A polytope is just a generalization of the higher-dimensional analogue of a polygon or a polyhedron. We'll start with some basic notions.

Definition 4.1. A set $S \subseteq \mathrm{R}^d$ is called *convex* if for any two points $x, y \in S$, every point $tx + (1-t)y$, for $0 \le t \le 1$, is also contained in S. (That is, S contains the whole line segment from x to y.)

Exercise 4.1. Let $\{C_i\}_{i2I}$ be a collection of convex subsets of R^d. Show that $\bigcap_{i2I} C_i$ is convex. (Note: I needn't be countable!)

Definition 4.2. A point $y \in \mathrm{R}^d$ is a *convex combination* of the points $x_1, \dots, x_n \in \mathrm{R}^d$ if there are real numbers $\alpha_1, \dots, \alpha_n\ in [0,1]$ such that $\sum_{i=1}^n \alpha_i = 1$ and $y = \alpha_1 x_1 + \cdots \alpha_n x_n$. The set of all convex combinations of a point set X, called the *convex hull* of X, is denoted $\mathrm{conv}(X)$.

For example, the unit ball in R^d is convex, but the unit sphere is not. Now we can introduce polytopes.

Definition 4.3. A *polytope* is the convex hull of a finite set of points.

It turns out, though, that polytopes have an alternate description.

A *hyperplane* in R^d is an a□ne subspace with dimension $d-1$. In other words, it's a translation of a $(d-1)$-dimensional linear subspace. For any $(d-1)$-dimensional subspace L of R^d, we choose a vector $u \in \mathrm{R}^d$ that is orthogonal to L; then the subspace can be written as $L = \{v \in \mathrm{R}^d : \langle u, v\rangle = 0\}$. Given any vector $w \in \mathrm{R}^d$, the shift $H = L + w$ is an a□ne subpsace of R^d. Moreover, its elements can be written as $\{v \in \mathrm{R}^d : \langle u, v\rangle = \langle u, w\rangle\}$. This description of hyperplanes is often more useful in practice.

Definition 4.4. Let $u \in \mathrm{R}^d$ and $c \in \mathrm{R}$. The *hyperplane* determined by u and c is

$$H_u(c) = \{v \in \mathrm{R}^d : \langle u, v\rangle = c\}.$$

The *(closed) positive half space* associated to $H_u(c)$ is the set

$$H_u^(c) = \{v \in \mathrm{R}^d : \langle u, v\rangle \ge c\}.$$

The negative half-space reverses the inequality, and the corresponding open half-spaces make the inequality strict.

In short, a half-space is the set of point on one side of the hyperplane.

Proposition 4.5. *A set $P \subseteq \mathbb{R}^d$ is a polytope if and only if it is a bounded intersection of finite number of closed half-spaces.*

The idea between the correspondence is that polygons and polyhedra can be characterized in two ways: By their vertices, corresponding to the convex hull, or by their faces, corresponding to the half-spaces. (Think for a moment about how a triangle, for example, is the intersection of three half-spaces.) Proposition 4.5 indicates that this is the case for higher-dimensional polytopes, as well.

Since each positive half-space is just a linear inequality, this means that polytopes are exactly the sets that can be represented by a finite list of linear inequalitites:

$$\begin{aligned} a_{1,1}x_1 + \cdots + a_{1,d}x_d &\le c_1 \\ a_{2,1}x_1 + \cdots + a_{2,d}x_d &\le c_2 \\ \vdots \qquad\qquad\quad \vdots\quad &\quad \vdots \\ a_{n,1}x_1 + \cdots + a_{n,d}x_d &\le c_n, \end{aligned}$$

or simply the matrix $A = (a_{i,j})$.

Exercise 4.2. A *minimal generating set* of hyperplanes for a polytope P is a set of hyperplanes $H_1, \ldots, H_n$ such that $P = \bigcap_{i=1}^n H_i^+$ and $\bigcap_{i \neq j} H_i^+$ strictly contains P for every $j \in [n]$. Suppose that $P = \bigcap_{i=1}^m H_i^+$.

1. Show that $\{H_1, \ldots, H_m\}$ contains a minimal generating set for P.
2. Show that this subset is unique.
3. Show that any two minimal generating sets for P are equal.

4.2. faces, f-vectors, and h-vectors

Definition 4.6. The *dimension* of a polytope P is d if P is contained in a d-dimensional affine subspace but not in any $(d-1)$-dimensional affine subspaces.

Any polyhedron can be broken up into smaller pieces with varying dimensions: faces, edges, and vertices. Our next goal is to extend this to polytopes.

Definition 4.7. Let $u \in \mathbb{R}^d$ be a vector and P a polytope. The *support* of P in the direction u is $\operatorname{supp}_u(P) = \max_{x \in P} \langle x, u \rangle$.

The hyperplane $H_u\big(\operatorname{supp}_u(P)\big)$ is the tangent hyperplane to P whose closed negative half-space contains P.

Definition 4.8. The *supporting face* of a polytope P in the direction of $u \in S^{d-1}$ is the set $\{x \in P : \langle x, u \rangle = \operatorname{supp}_u(P)\}$.

Note that every supporting face of P is itself a polytope of dimension at most one less than P. The set of *faces* of P is, well, the collection of its supporting faces. Faces with dimension 0, 1, and $\dim(P) - 1$ are called *vertices*, *edges*, and *facets*, respectively.

Definition 4.9. Suppose that P is a polytope with dimension d. The vector $(f_0, \ldots, f_d)$ in which f_k is the number of k-dimensional faces of P is called the *f-vector* of P. The f-polynomial of P is $f_P(t) = \sum_{k=0}^d f_d t^d$.

In other words, $f_P(t)$ is the generating function for the dimension statistic on the set of faces of P.

Example 4.10. The cube has 8 zero-dimensional faces (vertices), 12 one-dimensional faces (edges), 6 two-dimensional faces (facets), and 1 three-dimensional face. So its f-vector is $(8, 12, 6, 1)$.

Under certain conditions, the f-vector can be transformed into something particularly nice.

Example 4.11. The face generating function of the cube is $f(t) = 8 + 12t + 6t^2 + 1$. If we set $h(t) = f(t-1)$, then $h(t) = 1 + 3t + 3t^2 + 1$.

Example 4.12. The face generating function of the simplex is $f(t) = 4 + 6t + 4t^2 + t^3$, and $f(t-1) = 1 + t + t^2 + t^3$.

In our case, "particularly nice" means this:

Definition 4.13. A d-dimensional polytope P is *simple* if every vertex is incident to exactly d edges. In this case, the *h-vector* of P is $(h_0, \ldots, h_d)$, where $\sum_{i=0}^{d} h_i t^i = f(t-1)$.

These polytopes are simple in the respect that any vertex of a d-dimensional polytope must be incident to at least d edges. But it's also possible to prove that a polytope is simple if and only if its face structure (that is, the poset of its faces ordered by inclusion) is unchanged by a small translation of any of one its faces.

Here's the punchline: For simple polytopes, the h-vector is remarkably simple.

Theorem 4.14. *If P is a simple d-dimensional polytope and $(h_0, \ldots, h_d)$ is its h-vector, then*

1. $h_i \geq 0$ *for every* $0 \leq i \leq d$*,*
2. $h_i = h_{d-i}$ *for every* $0 \leq i \leq d$*, and*
3. $h_0 \leq h_1 \leq \cdots \leq h_{\lfloor d/2 \rfloor}$.

We will prove parts (1) and (2). The equalities in part (2) are called the DehnSommerville equations. Part (3), that the h-vector is unimodal, is remarkably difficult; Stanley first proved it by relating the h-vector to toric varieties and Betti numbers. There are less difficult proofs now, but none of them are simple.

Parts (1) and (2) follow from a theorem that provides a visual intuition for the h-vector (Theorem 4.17).

First, some setup:

Definition 4.15. A linear function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is called *generic* with respect to a polytope P if it is not constant on any edge.

Each linear function is given by a vector in $\mathbb{R}^d$. Each edge only disallows a subspace of these vectors with dimension $d-1$, so a generic function always exists.

Definition 4.16. The *1-skeleton* of a polytope P is the graph comprised of the vertices and edges of P.

Theorem 4.17. *Let G be the 1-skeleton of a polytope P and φ be a generic linear function with respect to P. If $\vec{G}$ denotes the graph obtained by orienting each edge $\{u, v\}$ from u to v if $\varphi(u) < \varphi(v)$, then h_i is equal to the number of vertices in $\vec{G}$ with indegree i.*

Proof. For each face F in P, let v_F denote the vertex in F where α is maximized. Then

$$f(t) = \sum_{F \subseteq P} t^{\dim(F)} = \sum_{v \in V(G)} \sum_{F : v_F = v} t^{\dim(F)}.$$

Since P is simple, every subset of the incoming edges of v in $\vec{G}$ determines a face F for which $v_F = v$; moreover, every face is determined by such a subset. So

$$\sum_{F : v_F = v} t^{\dim(F)} = \sum_{S \subseteq \left\{\substack{\text{incoming} \\ \text{edges}}\right\}} t^{|S|} = (1+t)^{\operatorname{indeg}(v)}.$$

Therefore $f(t) = \sum_{v \in V(G)} (1+t)^{\operatorname{indeg}(v)}$, and plugging in $h(t) = f(t-1)$ finishes the proof. ∎

4.3. the permutohedron

Definition 4.18. The *permutohedron* Π_n is the convex hull of the points $\{(\sigma_1, \ldots, \sigma_n) : \sigma \in S_n\} \subseteq F^n$.

The permutohedron is at most $(n-1)$-dimensional, since $\sigma_1 + \cdots + \sigma_n = \binom{n+1}{2}$ for every $\sigma \in S_n$. On the other hand, it is not contained in any $(n-2)$-dimensional a□ne subspace.

Exercise 4.3. Show that Π_n contains a segment in the direction of $e_1 - e_j$ for every $j \in \{2, 3, \ldots, n\}$. Then show that Π_n is not contained in any $(n-2)$-dimensional a□ne subspace of R^n.

Proposition 4.19. *Π_n has $n!$ vertices.*

Since Π_n is defined as the convex hull of $n!$ points, it has at most this many vertices. This proposition shows that, in fact, every one of these points is a vertex.

It's possible to show directly that none of the points $(\sigma_1, \ldots, \sigma_n)$ is in the convex hull of the others. You can also directly write down a linear function which is maximized at a given vertex.

Exercise 4.4. Do so.

Here's another way: Any polytope has at least one vertex, so Π_n has a vertex among the $n!$ vertices whose convex hull generate it. But Π_n is also invariant under the action of the symmetric group, so every image of this point must be a vertex; so Π_n has $n!$ vertices.

Let's find the supporting faces of Π_n. Let $\alpha \in \mathrm{R}^n$. First:

Exercise 4.5. Show that, if $\alpha_1 < \cdots < \alpha_n$, then $(1, 2, \ldots, n)$ is the sole maximizing point in Π_n of the function $\alpha(x) = \sum_{i=1}^n \alpha_i x_i$. Then show that if $\alpha_{\sigma(1)} < \cdots < \alpha_{\sigma(n)}$, then $\left(\sigma^{-1}(1), \ldots, \sigma^{-1}(n)\right)$ is the sole maximizing point of $\alpha(x)$ in Π_n.

If $\alpha_1 = \alpha_2 < \alpha_3 < \cdots < \alpha_n$, then two vertices of Π_n, specifically $(1, 2, \ldots, n)$ and $(2, 1, 3, 4, \ldots, n)$ maximize $\alpha(x)$, so the corresponding face is an edge. In the general case, we have

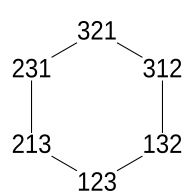
$$\alpha_{i_1} = \cdots = \alpha_{i_{n_1}} < \alpha_{i_{n_1+1}} = \cdots = \alpha_{i_{n_2}} < \cdots < \alpha_{n_{k-1}+1} = \cdots = \alpha_{n_k}.$$

In this case, the face that maximizes α is equal to the convex hull of the vertices where the numbers $\{n_{j-1} + 1, \ldots, n_j\}$ fill the coordinates in positions $i_{n_{j-1}+1}, \ldots, i_{n_j}$ (we set $n_0 = 0$ for convenience). So each face is congruent to $S_{I_1} \times S_{I_2} \times \cdots \times S_{I_k}$, where $I_j = \{n_{j-1}, \ldots, n_j\}$ and $n_0 = 0$: a product of smaller-dimensional permutohedra!

Moreover, these faces of Π_n are indexed by ordered set partitions, and an ordered set partition with k blocks corresponds to a face of dimension $n - k$. Since there are $k!\, S(n, k)$ ordered set partitions with k elements, the face vector of Π_n is given by

$$f_k = (n-k)!\, S(n, n-k).$$

In particular, each edge of Π_n is determined by an ordered set partition with $n-1$ blocks. If i and j share a block, then the vertices $(\sigma_1, \ldots, i, \ldots, j, \ldots, \sigma_n)$ and $\sigma_1, \ldots, j, \ldots, i, \ldots, \sigma_n)$ share an edge. In other words, the edges of Π_n are generated by the adjacent transpositions $(i\ \ i+1)$. Here is a picture of Π_3:



To find the h-vector of the permutohedron, pick some $\alpha \in \mathbb{R}^n$ with $\alpha_1 > \alpha_2 > \cdots > \alpha_n$ and direct the edges of Π_n according to its value. We know that each edge of Π_n corresponds to an adjacent transposition. So pick two vertices $\sigma = (\sigma_1, \dots, \sigma_n)$ and

$$\tau = (\sigma_1, \dots, \sigma_{i-1}, \sigma_j, \sigma_{i+1}, \dots, \sigma_{j-1}, \sigma_i, \sigma_{j+1}, \dots, \sigma_n).$$

These two vertices are connected by an edge directed from σ to τ if $i < j$ and $\sigma_i = \sigma_j - 1$. So

$$\operatorname{indeg}(\sigma) = \#\{\ell \in \{1, 2, \dots, n-1\} : \ell + 1 \text{ is to the left of } \ell \text{ in } \sigma\}.$$

Definition 4.20. We say that there is a *descent* at i in a permutation $(\sigma_1, \dots, \sigma_n)$ if $\sigma_i > \sigma_{i+1}$. The number of descents in a permutation σ is denoted $\operatorname{des}(\sigma)$.

Exercise 4.6. Let $\sigma \in S_n$. Why is $0 \le \operatorname{des}(\sigma) \le n-1$? Prove that $(-1)^{\operatorname{des}(\sigma)}$ is equal to the *sign* of σ (which is 1 if σ can be written as an even number of transpositions and -1 otherwise).

The indegree of σ counts something that's not exactly a descent. Actually, it counts the number of descents in the inverse permutation σ^{-1}.

Exercise 4.7. Show that $\operatorname{indeg}(\sigma) = \operatorname{des}(\sigma^{-1})$. (**Hint:** Writing σ in two-line notation may help.)

Definition 4.21. The *Eulerian number* $A(n,k)$ is the number of permutations in S_n with exactly k descents.

⚠ The Eulerian numbers are sometimes denoted $E(n,k)$ or $\left\langle {n \atop k} \right\rangle$. And sometimes $A(n,k)$ (or any of these other symbols) denotes the number of permutations with $k-1$ descents (meaning permutations with k "runs" of increasing numbers).

So we've determined the h-vector:

Proposition 4.22. *The coordinates of the h-vector of Π_n are given by $h_k = A(n,k)$.*

The relation $f(t) = h(t+1)$ gives us the following result:

Corollary 4.23. *For every $n \in \mathbb{N}$,*

$$\sum_{k=0}^{n-1} (n-k)!\, S(n, n-k) t^k = \sum_{k=0}^{n-1} A(n,k)(t+1)^k$$

Problem 4.8. Find a bijective proof of Corollary 4.23.

4.4. eulerian numbers and the γ-vector

First, some facts about the Eulerian numbers:

Exercise 4.9. Show that $A(n,k) = A(n, n-k-1)$. Then show that

$$A(n,k) = (n-k)A(n-1, k-1) + (k+1)A(n-1, k).$$

(**Hint:** What do you get if you remove the letter n from a permutation in S_n?)

We can arrange these numbers in a triangle, just like the binomial coefficients:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 4 & & 1 & & & \\
 & & 1 & & 11 & & 11 & & 1 & & \\
 & 1 & & 26 & & 66 & & 26 & & 1 & \\
1 & & 57 & & 302 & & 302 & & 57 & & 1
\end{array}$$

In fact, Pascal's triangle represents the h-vector of the n-cube.

Exercise 4.10. Show that the h-vector of the n-cube $[0,1]^n$ has coordinates $h_k = \binom{n}{k}$.

There's a stronger connection between these two triangles, though. Every row in the Eulerian triangle can be written as an integer linear combination of the rows of Pascal's triangle. For example, if we let P_n and E_n denote the nth rows of Pascal's and the Eulerian triangles, respectively, then

$$\begin{aligned} E_1 &= P_1 \\ E_2 &= P_2 \\ E_3 &= P_3 + 2P_1 \\ E_4 &= P_4 + 8P_2 \\ E_5 &= P_5 + 21P_3 + 14P_1 \\ &\vdots \end{aligned}$$

Actually, this is not so unusual: Any triangular array in which the first and last entry of each row is 1 will form a basis for the space of finite symmetric integer sequences. (Simply the 1 at each end allows you to always pick the next coefficient unobstructed.) Here's the special part: The coefficients in the linear expression of E_n from the rows P_n are always *positive*. These coefficients comprise the so-called γ-vector of Π_n.

Definition 4.24. If P is a simple polytope, its *γ-vector* is has the form $(\gamma_0, \ldots, \gamma_{\lfloor d/2 \rfloor})$ with coefficients given by

$$h_P(t) = \sum_{i=0}^{\lfloor d/2 \rfloor} \gamma_i t^i (t+1)^{d-2i}.$$

The γ-polynomial of P is $\gamma(t) = \sum_{i=0}^{\lfloor d/2 \rfloor} \gamma_i t^i$.

The term $(t+1)^{d-2i}$ expands out to the $(d-2i)$th row of Pascal's triangle, and the factor of t^i is included to "center" the row. The h- and f-polynomials are related by the simple expression $h(t) = f(t-1)$; the h- and γ-polynomials are related by the not-quite-so-simple-but-not-quite-complicated expression

$$h(t) = (1+t)^d \, \gamma\Big(\frac{t}{(1+t)^2}\Big).$$

Anyway, here's the surprising result, stated using this vector:

Proposition 4.25. *The γ-vector of Π_n is nonnegative for every $n \in \mathbb{N}$.*

This is encompassed by a much larger conjecture:

Conjecture 4.26 (Gal, 2005). *The γ-vector of any flag simple polytope has nonnegative entries.*

So what's a flag polytope?

Definition 4.27. Let P be a simple polytope. If $\bigcap F$ is nonempty for every collection $\mathcal{F}$ of faces of P whose elements all pairwise intersect, then P is called a *flag polytope*.

There's a different characterization, but we need a new concept. Feel free to skim this or skip it; the next section doesn't use it at all.

The *dual* of a polytope P is another polytope P^* whose face structure is the inverse of P. (More technically, the face *poset* of P^* is the dual of the face poset of P.) In other words, you can map the k-dimensional faces of P to the $(d-k)$-dimensional faces of P^* so that the inclusion ordering is reversed.

For example, the standard octahedron $\operatorname{conv}(\pm e_1, \pm e_2, \pm e_3)$ is the dual of the cube $[0,1]^3$. We map each face of the cube to the vertex of octahedron that it contains, each vertex to the face of the octahedron in the same orthant, and each edge connecting two faces in $[0,1]^3$ to the edge that connects the corresponding vertices in the octahedron. While the cube has 8 vertices, each of which is connected to 3 edges, each of which is, in turn, a member of 2 faces; the octahedron has 8 faces, each of which contains 3 edges, each of which, in turn, contains two vertices.

There is actually an explicit operation taking any convex body to its dual; the operation is not as *ad hoc* as the previous description might make it seem. For us, right now, the actual definition isn't so important; the crucial property is that it reverses the face structure of a polytope.[1]

If P is a simple polytope, then P is a simplicial complex. (In case you haven't yet met this definition, a *simplicial complex* on a set E is a nonempty collection $\mathcal{S}$ of subsets of E such that whenever $T \subseteq S$ and $S \in \mathcal{S}$, also $T \in \mathcal{S}$. An element of $\mathcal{S}$ is called a *face*.) The alternative characterization is that P is a flag polytope when P is a *clique simplicial complex*; that is, F is a face of P if and only if the vertices it contains form a clique in the 1-skeleton of P.

Question 4.28. What is a combinatorial interpretation of the γ-vector?

4.5. volume of polytopes

When their namesake studied the Eulerian numbers, he did it in the context of polynomials. Consider this sequence:

$$\begin{aligned} x + x^2 + x^3 + \cdots &= \frac{x}{1-x} \\ x + 2x^2 + 3x^3 + \cdots &= \frac{x}{(1-x)^2} \\ x + 4x^2 + 9x^3 + \cdots &= \frac{x}{(1-x)^3}(1+x) \\ x + 8x^2 + 27x^3 + \cdots &= \frac{x}{(1-x)^4}(1+4x+x^2) \end{aligned}$$

Parts of these polynomials look a lot like generating functions for the Eulerian numbers. In fact, Euler took this as the *definition* of these numbers:

$$\sum_{k=1}^{\infty} k^n x^i = \frac{x}{(1-x)^{n+1}} A_n(x).$$

We'll prove it as a theorem.

Theorem 4.29. $A_n(x) = \sum_{k=0}^{n\ 1} A(n,k) x^k$.

The first thing we need to check, though, is that there even *is* a polynomial that we can multiply by $x(1-x)^{(n+1)}$ to get the generating function for $(k^n)_{k=1}^{\infty}$? Actually, it's not so hard: we know that

$$\left(x \frac{d}{dx}\right)^n \frac{x}{1-x} = \sum_{k=1}^{\infty} k^n x^k.$$

Expanding out the left side is a mess, but you can check that the result is a big sum of polynomials over the denominator $(1-x)^n$. In fact, you can prove Theorem 4.29 this way, using induction.

[1] If P $\mathbb{R}^d$ is a convex body, then

$$P^\circ := \{ y \in \mathbb{R}^d : \langle x, y \rangle \ 1 \}.$$

Each point of P° corresponds to a closed half-space that contains both the origin and P (and vice-versa). For this reason, the polar is usually only applied to convex bodies that contain the origin; this is usually no problem, since we can translate it to ensure that $0 \in P$.

Once you expand everything out, the proof essentially comes down to the recurrence relation for the Eulerian numbers. But that proof is neither fun nor enlightening, so let's do something else.

The volume of a polytope is exactly what you would expect: The integral of 1 over its region in space. (Or, if you prefer, the Lebesgue measure of the polytope.) The volume of a polyhedron (the not-necessarily-bounded intersection of half-spaces) P in the positive orthant $\mathbb{R}^d_{\geq 0}$ may not be finite, but we can still get a handle on it by stratifying by slices. We'll use hyperplanes with points $\sum_i x_i = k$ for some $k \in \mathbb{Z}$. Let $\mathrm{Sl_k}$ denote the kth "slice," that is,

$$\mathrm{Sl}_k = \{\mathbf{x} \in \mathbb{R}^d_{\geq 0} : k-1 \leq x_1 + \cdots + x_d \leq k\}$$

and $\mathrm{Sl}_k(P)$ denote the region $\mathrm{Sl}_k \cap P$. Since Sl_k is a bounded region, the volume of $\mathrm{Sl}_k(P)$ is finite. So for any polyhedron $P \subseteq \mathbb{R}^d_{\geq 0}$, we can define its slice-volume generating function

$$S(P) = \sum_{k=1}^{\infty} d! \operatorname{Vol}\big(\mathrm{Sl}_k(P)\big) x^k.$$

What's the $d!$ for? Well, it turns out that if P is a *lattice polytope* (each vertex is a point in $\mathbb{Z}^d$), then the volume of P is a rational number whose denominator divides $d!$, so $d! \operatorname{Vol}(P)$ is an integer for every lattice polytope P.

It turns out that the Eulerian numbers already have an interpretation in terms of this slice polynomial.

Theorem 4.30. $S([0,1]^n) = xA_n(x)$.

Apparently this result is incipient in some 1886 work of Laplace, though it has a completely different focus. Richard Stanley provided a one-page proof of Theorem 4.30 in a 1977 paper.

Problem 4.11. Read and understand this paper.

The slices $\mathrm{Sl}_k([0,1]^n)$ are called *hypersimplices*, denoted by $\Delta_{n,k}$;[2] so $n! \operatorname{Vol}(\Delta_{n,k}) = A(n, k-1)$. The special case $\Delta_{n,1}$ is usually denoted simply by Δ_n. It is the convex hull of $e_1, \ldots, e_n$ and the origin.

We're almost ready to prove Theorem 4.29. We just need one more ingredient.

Lemma 4.31. $\operatorname{Vol}(\Delta_n) = \frac{1}{n!}$.

Proof. The case $n = 1$ is clear. For $n \geq 1$, we stratify Δ_n by slices parallel to the nth coordinate; thus

$$\operatorname{Vol}_n(\Delta_n) = \int_0^1 t^{n-1} \operatorname{Vol}_{n-1}(\Delta_{n-1})\, dt = \frac{1}{n} \operatorname{Vol}_{n-1}(\Delta_{n-1}),$$

which is $\frac{1}{n}$ by induction. □

Corollary 4.32. $n! \operatorname{Vol}(\mathrm{Sl}_k) = k^n - (k-1)^n$.

Proof. $\mathrm{Sl}_k = k\Delta_n \setminus (k-1)\Delta_n$ (up to a set of measure 0). □

Let's do it.

Proof of Theorem 4.29. From Corollary 4.32, we see that $S(\mathbb{R}^n_{\geq 0}) = (1-x)\sum_{k=1}^{\infty} k^n x^k$; combining this with Theorem 4.30, it suffices to prove that

$$S(\mathbb{R}^n_{\geq 0}) = \frac{S([0,1]^n)}{(1-x)^n}.$$

[2] An alternative definition of $\Delta_{n,k}$ is the convex hull of the vectors in $\{0,1\}^n$ that have exactly k ones and $n-k$ zeros. The polytope defined in this way is geometrically different but combinatorially the same.

Now, $S(P+e_i) = xS(P)$ for every standard basis vector e_i. In particular, the unit cube tiles the positive orthant, so

$$\mathsf{R}^n{}_0 = [0,1]^n \sum_{i=1}^{n} (1 + e_i + 2e_i + \cdots),$$

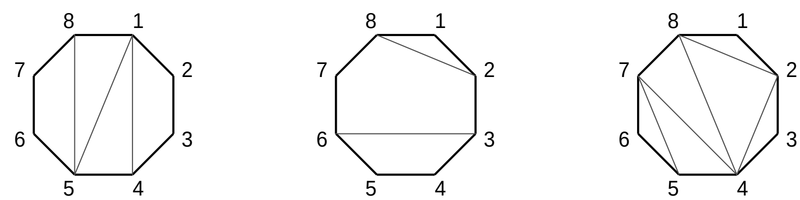
so

$$S(\mathsf{R}^n{}_0) = S([0,1]^n) \prod_{i=1}^{n} (1 + x + x^2 + \cdots) = \frac{S([0,1]^n)}{(1-x)^n}. \qquad \square$$

Question 4.33. Can you find a bijective proof of Theorem 4.29?

4.6. the associahedron

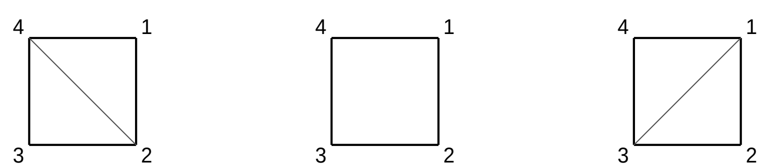
A *noncrossing subdivision* of a polygon is what it sounds like: A set of noncrossing lines that connect vertices of a polygon, dividing it into several pieces. For example, here are three different subdivisions of a regular octagon:



The rightmost subdivision is called a *triangulation*, simply because the polygon is divided into triangles. A noncrossing subdivision is a triangulation if and only if no further noncrossing lines can be added.

Definition 4.34. The n-dimensional *associahedron* Assoc_n is an "abstract polytope"—that is, a poset of faces—consisting of the noncrossing subdivisions of the regular $(n+2)$-gon, ordered by reverse refinement. (If α refines β, then $\alpha < \beta$.)

In other words, Assoc_n is like the face structure of a polytope divorced from any geometrical rendering. In fact, it's not obvious that it *has* a geometric realization, but we'll get to that. For now, consider Assoc_2, the subdivisions of a square. There are three of them:

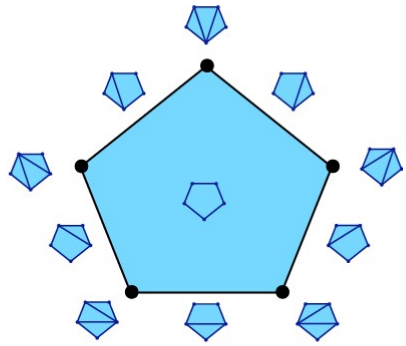


The left and the right subdivision refine the middle one, so we think of the middle subdivision as representing a face that contains the two other subdivisions, each of which is a vertex.

Here is a representation of Assoc_3:[3]

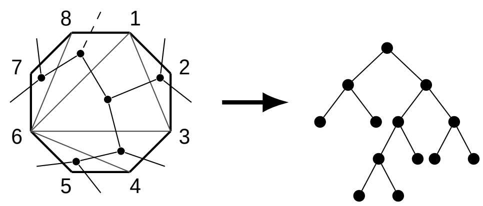
---

[3] This is not very visually appealing, but I didn't want to spend the time trying to figure out how to place Ti*k*Zpictures inside nodes of a meta-Ti*k*Zpicture. This graphic is from this website.



Exercise 4.12. Show that the number of triangulations of a regular $(n+2)$-gon is the Catalan number C_n. (In other words, Assoc_n has C_n vertices.)

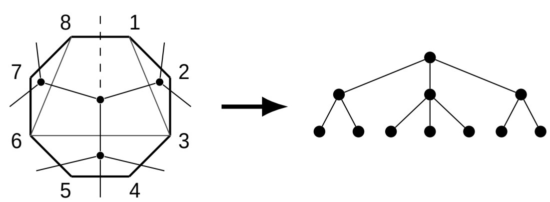
What about the number of k-dimensional faces? (That is, the number of noncrossing subdivisions of an $(n+2)$-gon with $n-1-k$ lines.) Here's one way to count them. For clarity, we'll start with 0-dimensional faces: triangulations. Given a triangulation, we can form a complete binary tree out of it, like this:



The process for making this tree is to first place a point in every triangle, then draw a line through every edge between adjacent triangles, and also a line through each boundary edge. Deleting the line that passes through the edge between vertex 1 and vertex $n+2$ produces a complete binary tree. (When drawing the binary tree on the right, it's important to keep the relative locations of the child and parent nodes rotationally the same.)

Exercise 4.13. Show that this is a bijection from the set of triangulations of an $(n+2)$-gon to the set of complete binary trees with $n+1$ vertices.

You can pull the same stunt for noncrossing subdivisions that aren't necessarily triangulations, For example:



Each face of Assoc_n corresponds to a plane tree with $n+1$ leaves and the additional property that every non-leaf vertex has at least two children. A face of dimension k corresponds to such a tree with $n-k$ non-leaf vertices.

Now we come the the crucial question: Is the associahedron an actual polytope? In other words, can the face structure of Assoc_n be *geometrically realized*? The answer is yes, and Jean-Louis Loday described a particularly nice one in 2004; explaining it and proving that it does, in fact, realize the associahedron comprise the next section.

4.7. geometric realization of the associahedron

Definition 4.35. Given two set $X, Y \in \mathbb{R}^n$, their *Minkowski sum* is

$$X + Y = \{x + y : x \in X \text{ and } y \in Y\}.$$

With this simple definition, we can define the polytope.

Definition 4.36. Let $e_1, \ldots, e_n$ denote the standard basis in $\mathbb{R}^n$. For each $1 \leq a \leq b \leq n$, define $\Delta_{[a,b]} := \operatorname{conv}(e_a, e_{a+1}, \ldots, e_b)$. Then

$$A_n := \sum_{1 \leq a \leq b \leq n} \Delta_{[a,b]},$$

where the symbol $\sum$ denotes the Minkowski sum.

Theorem 4.37. *The polytope A_n is a geometric realization of Assoc_n.*

To prove it, we'll need a lemma. Recall that $\operatorname{supp}_u(P)$ is the support of u on P: The set of points in P on which $\langle u, x\rangle$ is maximized.

Lemma 4.38. *If P and Q are polytopes in $\mathbb{R}^n$, then*

$$\operatorname{supp}_u(P + Q) = \operatorname{supp}_u(P) + \operatorname{supp}_u(Q).$$

Proof. Pick any $u \in \mathbb{R}^n$. Every point in $P + Q$ has the form $p + q$ with $p \in P$ and $q \in Q$. If $\langle p, u\rangle$ and $\langle q, u\rangle$ are maximal, then their sum $\langle p + q, u\rangle$ must be, as well. Conversely, if one of the inner products is not maximal, then $\langle p + q, u\rangle$ is not maximal, either. □

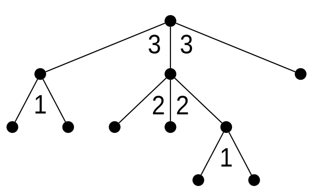
On to the proof.

Most of a proof of Theorem 4.37. Fix some $u \in \mathbb{R}$. Lemma 4.38 indicates we should consider $\operatorname{supp}_u(\Delta_{[a,b]})$. But this is easy:

$$\operatorname{supp}_u(\Delta_{[a,b]}) = \operatorname{conv}\left(e_k : a \leq k \leq b \text{ and } u_k = \max_{a \leq i \leq b} u_i\right).$$

In particular, supp_u depends only on the relative order of the coordinates of u.

Now we show a bijection between the faces of A_n and plane trees with $n + 1$ vertices in which every non-leaf vertex has at least two children. Given a sequence $(u_1, \ldots, u_n)$, we form the tree iteratively: If largest value in the sequence occurs k times, then form a root node with $k + 1$ children. Removing the instances of the largest value breaks u into $k + 1$ subsequences (some of them possibly empty); at the ith child vertex, repeat this process with the ith subsequence. For example, if $u = (1, 3, 2, 2, 1, 3)$, then



As the labels indicate, each "nook" in the tree corresponds to a coordinate of u. Now we want to ask: If u and v produce the same plane tree, does $\operatorname{supp}_u(A_n) = \operatorname{supp}_v(A_n)$? And what about the converse?

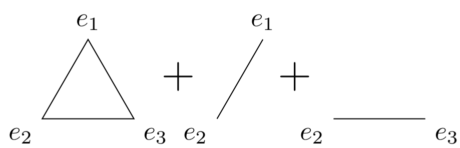
The answer is affirmative for both, because the plane tree determines the support on $\Delta_{[a,b]}$ and vice versa. To determine the support on $\Delta_{[a,b]}$, simply consider the highest nook(s) that are between the ath and $(b + 1)$th leaves. The corresponding coordinate(s) correspond to the vertices in the support—no matter the vector u that generates the tree. For example, for the tree above, $\mathrm{supp}_u(\Delta_{[3,5]}) = \mathrm{conv}(e_3, e_4)$, since the highest nooks correspond to the third and fourth coordinates. Conversely, the support on the sets $\Delta_{[a,b]}$ completely determines the "nook structure" of the tree.

A face $F = \mathrm{supp}_u(A_n)$ is contained in a second face $G = \mathrm{supp}_w(A_n)$ when one or more of the inequalities between the coordinate values in u is replaced by equality in w. In the tree picture, this corresponds to retracting one or more edges. You may check that, in our bijection between noncrossing subdivisions and planar trees, erasing one of the edges in the subdivision corresponds to retracting an edge in the tree. So the face structure of A_n corresponds exactly to the face structure of Assoc_n. □

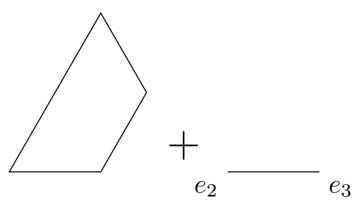
In fact, this proof gives a way to calculate the vertices of A_n.

Exercise 4.14. Let T be a complete binary tree T with $n+1$ leaves. For each $i \in [n]$, let v_i be the vertex at the ith nook of T. We define ℓ_i as the number of leaves in the left branch of v_i and r_v as the number of leaves in the right branch of v_i. Show that the vertices of A_n are the points $\mathbf{x}(T) = (\ell_1 r_1, \ell_2 r_2, \ldots, \ell_n r_n)$ as T varies across all complete binary trees with $n+1$ vertices.

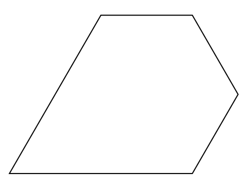
This particular realization of Assoc_n is not symmetric. Take, for example, the only one I can easily render here: A_3. It is the sum



It's missing the single vertices $\Delta_{[1,1]}$, $\Delta_{[2,2]}$, and $\Delta_{[3,3]}$, but these only translate the polytope, so we'll omit them. Adding the first two, we get



Adding these two together gives A_3:



‖ 4.8. graphical associahedra

We can use this realization of the associahedron to define a polytope associated to any graph.

Definition 4.39. Let G be a simple, undirected, connected graph G with vertex set $[n]$. A subset $I \subseteq [n]$ is called *G-connected* if $G|_I$ is connected; that is, if there is a path from any vertex in I to any other vertex in I using only edges that connect elements of I. The *graphical associahedron* associated to G is the polytope $A(G) \subseteq \mathsf{R}^n$ defined by

$$A(G) := \sum_{\substack{I \quad [n] \\ I \text{ is } G\text{-connected}}} \Delta_I,$$

where $\sum$ denotes the Minkowski sum and $\Delta_I = \mathrm{conv}(e_i : i \in I)$.[4]

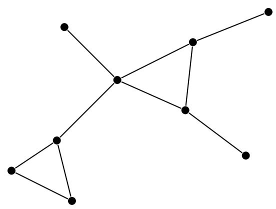
For example, when $G = P_n$, the path with n vertices, we get the associahedron A_n. Since $V(G)$ is always G-connected, $A(G)$ is a Minkowski sum of subsets of the $(n-1)$-dimensional simplex that includes the whole simplex. This means that $A(G)$ is an $(n-1)$-dimensional polytope.

Exercise 4.15. Show that $A(K_n)$ is a geometric realization of the permutohedron Π_n.

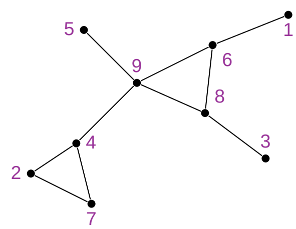
Just as before, we can calculate the faces of $A(G)$ as

$$\begin{aligned} \mathrm{supp}_u\big(A(G)\big) &= \sum_{\substack{I \subseteq [n] \\ I \text{ is } G\text{-connected}}} \mathrm{supp}_u(\Delta_I) \\ &= \sum_{\substack{I \subseteq [n] \\ I \text{ is } G\text{-connected}}} \Delta_{J(I)}, \end{aligned}$$

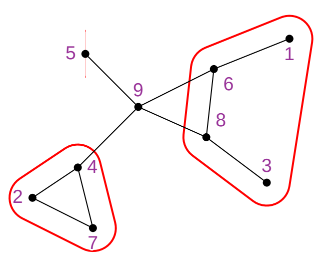
where $J(I) = \{j \in I : u_j = \max_{i \in I} u_i\}$. Let's begin by determining the vertices; we'll obtain those when each coordinate of the vector u is distinct. We'll do this by example. Suppose that G is this graph:



We can imagine the vector u by writing the value u_v at vertex v; perhaps our vector u looks like this:

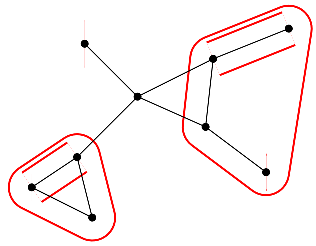


For the convenience of this example, we'll take the vertex label to be the same as its u-value, so $u_i = i$. If a G-connected set I contains u_9, then $\mathrm{supp}_u(I) = u_9$. If it doesn't contain 9, then I is completely contained in one of the three connected components of G that result after deleting vertex u_9:



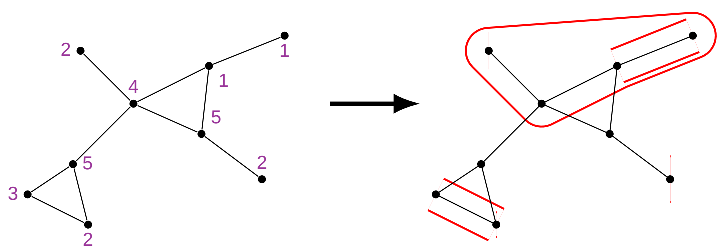
[4] These polytopes are also sometimes referred to as *nestohedra* for reasons that will become clear later.

If I is completely contained in the right component $\{u_1, u_3, u_6, u_8\}$, then $\mathrm{supp}_u(I) = u_8$—unless $u_8 \notin I$, in which case I is contained in either $\{u_1, u_6\}$ or $\{u_3\}$, the two components of $\{u_1, u_3, u_6, u_8\}$ obtained by deleting u_8. And so on. Repeating this process, we get a nested set of connected components like this:



To determine $\mathrm{supp}_u(I)$, simply take the smallest outlined region R that wholly contains I; then $\mathrm{supp}_u(I)$ is the vertex of R that is contained in the fewest regions overall.

This grouping of the vertices of G is all the information required to determine the vertex $\mathrm{supp}_u(A(G))$. This procedure can be extended to any vector u, even when some coordinates are equal. At each step, we draw regions that encompass the connected components once the highest u-valued vertices are deleted. For example:



Definition 4.40. A *tube* in a graph G is a nonempty proper subset of $V(G)$ which induces a connected subgraph. A *tubing* of G is a collection $\mathcal{T}$ of tubes such that, for every $I, J \in \mathcal{T}$,

1. either $I \subseteq J$, $J \subseteq I$, or $I \cap J = \emptyset$; and
2. if $I \cap J = \emptyset$, then $I \cup J$ is not a tube.

The second condition can be equivalently stated: If $I \cap J = \emptyset$, then there is no edge from a vertex in I to a vertex in J. If $\mathcal{T}$ is has k tubes, then we call it a *k-tubing*. We write $\mathcal{T} \subseteq \mathcal{T}'$ when every tube in $\mathcal{T}$ is a tube in $\mathcal{T}'$. (This is nothing more than usual set containment.) What this section hints at (and the proof is not much harder) is this:

Theorem 4.41. *Let G be a graph with n vertices. The set of faces in $A(G)$ of dimension k are in bijection with the set of $(n-1-k)$-tubings of G. Moreover, a face F contains another face F' if the tubing for F' contains the tubing for F.*

This theorem can be economically phrased in terms of a dual.

Definition 4.42. The *nested set complex* of a graph G, denoted $\mathcal{N}(G)$, is the simplicial complex on the set of tubes of G whose faces are the tubings of G.

Theorem 4.43. *For any graph G, the nested set complex $\mathcal{N}(G)$ is dual to the graphical associahedron $A(G)$. In other words, there is an inclusion-reversing bijection between the faces of $\mathcal{N}(G)$ and the faces of $A(G)$.*

This duality is the reason graphical associahedra are sometimes referred to as "nestohedra." The nested set complex has many nice properties: all the maximal subsets have the same cardinality (this property is called *pure*), it's a clique simplicial complex, and it's topologically equivalent to a sphere. For these polytopes, Gal's conjecture is not a conjecture; it's a theorem.

5. posets

5.1. definitions

Definition 5.1. A *partially ordered set*, or *poset*, is a set X together with a binary relation $\preccurlyeq$ such that

- $x \preccurlyeq x$ for every $x \in X$ (reflexivity);
- if $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$ (transitivity); and
- if $x \preccurlyeq y$ and $y \preccurlyeq x$, then $y = x$ (anti-symmetry).

The best-known example of a poset is the set of subsets of a given set S with the containment $\subseteq$ as the partial order. For a different example, say that $(x_1, \dots, x_n) \preccurlyeq (y_1, \dots, y_n)$ for two points $x, y \in \mathbb{R}^n$ if $x_i \le y_i$ for every $i \in [n]$. This, too, is a partial order.

For the rest of this section, $(X, \preccurlyeq)$ will always be a partially ordered set. Also, we use the notation $x \prec y$ to denote the fact that $x \preceq y$ and $x \neq y$.

Definition 5.2. We say that $y \in X$ *covers* $x \in X$ if $x \prec y$ and if $x \preceq z \preceq y$, then $z = x$ or $z = y$. In this case, we write $x \lessdot y$.

In other words, y covers x if there are are no elements strictly between them. If X is finite, then the covering relations completely determine the partial order. This information can be displayed visually in a *Hasse diagram*, where each element of X is a vertex, there is an edge from x to y if $x \lessdot y$, and the vertices are drawn so that all edges point upward. Infinite posets are not necessarily characterized by their covering relations. The poset $(\mathbb{Z}, \le)$ is, but $(\mathbb{Q}, \le)$ doesn't even have any covering relations.

Some posets have a special structure:

Definition 5.3. A *lattice* is a poset $(X, \preccurlyeq)$ such that each pair of elements x and y has a unique minimal upper bound, denoted $x \vee y$, and a unique maximal lower bound, denoted $x \wedge y$.

The poset $\big(\mathcal{P}(S), \subseteq\big)$ is a lattice: The minimal upper bound of U and V is $U \cup V$ and the maximal lower bound is $U \cap V$. (Note the similarity of $\cup$ to $\vee$ and $\cap$ to $\wedge$; this is in fact where the notation comes from.) This poset is called the *Boolean lattice* on the set S. The Boolean lattice on the set $[n]$ is denoted $\mathbb{B}_n$.

But there is an intrinsically algebraic definition of lattices, as well.

Definition 5.4. A set X with two binary operations $\wedge$ and $\vee$ is called an *algebraic lattice* if the two operations are both associative and commmutative and satisfy the *absorption laws*:

1. $x \vee (x \wedge y) = x$ and
2. $x \wedge (x \vee y) = x$

for every $x, y \in X$.

You can check that any Boolean lattice satisfies these absorption laws. If these operations really are analagous to set intersection and union, then we'd like them also to be idempotent. But it turns out that absorption already guaratees this:

Proposition 5.5. *If $(X, \wedge, \vee)$ is an algebraic lattice, then $x \wedge x = x$ and $x \vee x = x$.*

Proof. Applying the two absorption laws to the middle expression, we have

$$x \wedge x = x \wedge \big(x \vee (x \wedge x)\big) = x;$$

the other case is similar. □

While algebraic lattices seem a little off-kilter, it turns out are simply an algebraic description of poset lattices.

Proposition 5.6. *Every algebraic lattice is a poset lattice and vice versa.*

Proof sketch. To prove $(\Rightarrow)$, define $x \preccurlyeq y$ if and only if $x \vee y = y$. You can check that $\preccurlyeq$ is then a partial order and the resulting poset is a lattice. For $(\Leftarrow)$, define $\vee$ and $\wedge$ as the unique minimal upper and maximal lower bounds, respectively, and check the absorption laws. □

In posets, there is a distinction between *minimal* and *minimum* elements in a poset. An element x is called *minimal* if whenever $z \preccurlyeq x$, then $z = x$ (no element is smaller than x); it is called a *minimum* element if $x \preccurlyeq y$ for every element $y \in X$. For example, consider the set of nonempty subsets of $\{1, 2, 3, 4\}$ ordered by inclusion. This poset has 4 minimal elements—$\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$—but no minimum element. The same distinction can be made between *maximal* and *maximum* elements.

Exercise 5.1. Prove that, if a poset $(X, \preccurlyeq)$ has a minimum element, it is unique. (The same is true for maximum elements, of course.)

Because of this uniqueness, we use the notation $\hat{0}$ and $\hat{1}$ to denote the minimum and maximum elements of a poset—if they exist. In a finite lattice, they always do.

Proposition 5.7. *Every finite lattice contains a maximum and a minimum element.*

Proof. Because $\wedge$ is associative, the expresion $\bigvee_{x \in X} x$ is well-defined; it is an element that is less than or equal to every element of the lattice—so it is a minimum element. Similarly, $\bigwedge_{x \in X} x$ is a maximum element. □

Exercise 5.2. Find an infinite lattice with neither a minimum nor a maximum element. (Make sure it's a lattice, not just a poset!)

Definition 5.8. A *chain* in a poset $(X, \preccurlyeq)$ is a sequence of elements $(x_1, \dots, x_n)$ such that $x_1 \prec x_2 \prec \cdots \prec x_n$. A *saturated chain* is a chain that is not a proper subsequence of any other chain.

In B_3, the sequence $\emptyset \subset \{1\} \subset \{1, 2, 3\}$ is a chain, but it's not saturated because the chain $\emptyset \subset \{1\} \subset \{1, 2\} \subset \{1, 2, 3\}$ properly contains it.

Proposition 5.9. B_n *has exactly* $n!$ *saturated chains.*

Proof sketch. A saturated chain contains exactly one subset of size k for every $0 \le k \le n$. If $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_n = [n]$ is a chain, we can write down a sequence $(c_1, \dots, c_n)$, where c_i is the single element in $S_i \setminus S_{i-1}$. This sequence is necessarily a permutation, and this correspondence is a bijection between permutations and saturated chains. □

The *ordered Bell number* OB_n is the number of ordered set partitions of $[n]$. So $OB_n = \sum_{k=0}^{n} k! S(n, k)$. See Wikipedia for more information, including a variety of interesting alternate formulas.

Proposition 5.10. *The number of chains in* B_n *is* $4OB_n$.

Proof. If we can show that the number of chains that do not include $\emptyset$ or $[n]$ is OB_n, then we're done. To every chain $\emptyset \ne S_1 \subset S_2 \subset \cdots \subset S_k \ne [n]$, we associate the ordered set partition $(S_1, S_2 \setminus S_1, \dots, S_k \setminus S_{k-1}, [n] \setminus S_k)$. This is a bijection. □

What about relations between posets? It's much as you would expect.

Definition 5.11. If $(X, \preccurlyeq)$ and $(Y, \le)$ are posets, a map $f \colon X \to Y$ is called *order-preserving* if $f(x) \le f(y)$ whenever $x \le y$. We say that f is an *isomorphism* if f is an order-preserving bijection.

5.2. q-analogues and the lattice of subspaces

Now we'll look at a poset of a decidedly different kind. We use $\mathbb{F}_q$ to denote the finite field with q elements.[1]

Definition 5.12. If q is the power of a prime, we let $L_q(n)$ denote the poset of subspaces of $\mathbb{F}_q^n$, ordered by inclusion.

In fact, $L_q(n)$ is a lattice: If S and T are two subspaces of $\mathbb{F}_q^n$, then $S \cap T = S \wedge T$ and $\operatorname{span}(S \cup T) = S \vee T$. One interesting fact about $L_q(n)$ is that it is self-dual.

Definition 5.13. The *dual* of a poset $(X, \preccurlyeq)$ is the poset $(X, \preccurlyeq')$ where $x \preccurlyeq' y$ if and only if $y \preccurlyeq x$.

In other words, the dual of poset is obtained by reversing all the inequalities, or, more geometrically, by turning the Hasse diagram upside down.

Proposition 5.14. *The dual of $L_q(n)$ is isomorphic to $L_q(n)$.*

Exercise 5.3. Show that the map

$$S \mapsto \{x \in \mathbb{F}_q^n : \langle x, s \rangle = 0 \text{ for every } s \in S\}$$

is an isomorphism from $L_q(n)$ to its dual.

Our next goal is to count the number of saturated chains of $L_q(n)$. These are called the *complete flags* of $\mathbb{F}_q^n$. For each saturated chain $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_n = \mathbb{F}_q^n$, we can choose a basis $(e_1, \ldots, e_n)$ of $\mathbb{F}_q^n$ such that $(e_1, \ldots, e_k)$ is a basis for S_k; this can be done by simply choosing each $e_k \in S_k \setminus S_{k-1}$. It's easy to count the number of such sequences: The first vector e_1 can be any nonzero vector, so there are $q^n - 1$ options; the second vector can be any nonzero vector not in the span of e_1, so there are $q^n - q$ options; in general, there are $q^n - q^{k-1}$ options for e_k. So there are

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1})$$

sequences $(e_1, \ldots, e_n)$.

Each sequence determines a flag, but each flag is associated to multiple sequences. How many? If the flag $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_n = \mathbb{F}_q^n$ corresponds $(e_1, \ldots, e_n)$, then each e_k is in $S_k \setminus S_{k-1}$, so there are $q^k - q^{k-1}$ ways to choose e_k. In other words, each flag corresponds to

$$(q-1)(q^2-q) \cdots (q^n - q^{n-1})$$

different sequences. In total, then, there are

$$\frac{q^n-1}{q-1} \frac{q^n-q}{q^2-q} \cdots \frac{q^n-q^{n-1}}{q^n-q^{n-1}} = \frac{q^n-1}{q-1} \frac{q^{n-1}-1}{q-1} \cdots \frac{q^2-1}{q-1} \frac{q-1}{q-1}$$

saturated chains in $L_q(n)$.

In the remainder of the section, we'll connect this formula to the so-called *q-analogues*.[2] Roughly speaking, a q-analogue of a combinatorial object is a polynomial in q that specializes to that combinatorial object when $q = 1$. Of course, lots of polynomials will end up doing this, so it's a matter of experience to see what are the "useful" q-analogues. (See this blog post for a brief overview of the philosophy.)

We'll start with a q-analogue of the natural numbers.

---

[1] There is a finite field with q elements if and only if q is a power of a prime; if it exists, it is unique up to isomorphism. See Appendix A.1 for a very brief overview of finite fields, or see Keith Conrad's notes.

[2] Basically everywhere, this term is spelled q-analog. But I can't abide that—"analog" seems always to bring to mind the analog/digital dichotomy instead of the noun form of "analogous"—so I'm going to be contrarian throughout this section and include the *ue*.

Definition 5.15. The q-analogue of $n \in \mathbb{N}$ is the polynomial
$$[n]_q := 1 + q + \cdots + q^{n-1} = \frac{q^n - 1}{q - 1}.$$

One intuition for this definition is that it incorporates "how we count up to n" into its definition. But don't worry about that too much. We can rapidly extend our q-analogues to other combinatorial objects.

Definition 5.16. The q-factorial of n is
$$[n]_q! := [n]_q[n-1]_q \cdots [1]_q.$$

Actually, this q-factorial is an example of something that we've already seen:

Proposition 5.17. *$[n]_q!$ is the generating function for the number of inversions of permutations in S_n.*

(Given a permutation σ, an *inversion* of σ is a pair i, j with $1 \leq i < j \leq n$ such that $\sigma(i) > \sigma(j)$.) In other words, $[n]_q!$ is the generating function of a statistic. While $n!$ counts the number of permutations, $[n]_q!$ counts the number of partitions and stratifies them according to the number of inversions. You can see a little more of this interaction here.

We can use these q-factorials to rephrase our count of the complete flags in $L_q(n)$.

Proposition 5.18. *The number of saturated chains in $L_q(n)$ is $[n]_q!$.*

One last q-analogue:

Definition 5.19. The q-binomial coefficient is
$$\binom{n}{k}_q = \frac{[n]_q!}{[k]_q!\,[n-k]_q!}.$$

It seems like the q-binomial coefficient is unlikely to be a polynomial. Surprisingly, it always is. (Compare this to the corresponding definition $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. From this formula, it's not clear that $\binom{n}{k}$ is an integer.)

Exercise 5.4. By convention, we set $[0]_q! = 1$. Show that $\binom{n}{0}_q = \binom{n}{n}_q = 1$. Then show that $\binom{n}{1}_q = \binom{n}{n-1}_q = [n]_q$.

Exercise 5.5. Prove the recurrence relations
$$\binom{n}{k}_q = q^{n-k}\binom{n-1}{k-1}_q + \binom{n-1}{k}_q$$
and
$$\binom{n}{k}_q = q^{k}\binom{n-1}{k}_q + \binom{n-1}{k-1}_q.$$

Exercise 5.6. Use induction to show that $\binom{n}{k}_q$ is a polynomial with positive integer coefficients for every $n \in \mathbb{N}_0$ and $0 \leq k \leq n$.

Actually, the q-binomial coefficients have an interpretation in terms of $\mathbb{F}_q^n$, as well.

Proposition 5.20. *The number of subspaces of $\mathbb{F}_q^n$ of dimension k is $\binom{n}{k}_q$.*

Proof. For each subspace T of dimension k, there are $[k]_q![n-k]_q!$ different complete flags $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_n = \mathbb{F}_q^n$ such that $S_k = T$. (This is because $S_0 \subset \cdots \subset S_k$ can be any complete flag of the vector space S_k, which is isomorphic to $\mathbb{F}_q^k$, and the flags $S_k \subset \cdots \subset S_n$ are in bijection to the complete flags of the quotient vector space $\mathbb{F}_q^n/S_k$. These choices of flags are independent.)

Since there are $[n]_q!$ complete flags and each subspace of dimension k corresponds to exactly $[k]_q![n-k]_q!$ of them, there are $\binom{n}{k}_q$ subspaces of dimension k. $\square$

⚠ Be wary! Not everything generalizes directly to the world of q's. For example,

$$\sum_{k=0}^{n} \binom{n}{k}_q \neq \left([2]_q\right)^n = (1+q)^n.$$

Instead, the q-binomial formula is

$$\sum_{k=0}^{n} q^{k(k-1)/2} \binom{n}{k}_q = \prod_{k=0}^{n-1} (1+q^k).$$

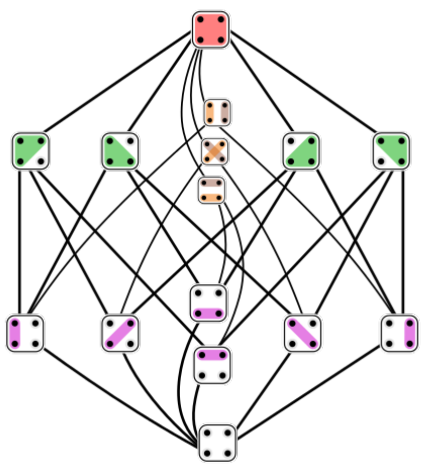
(One combinatorial proof of this appears on p. 74 of *Enumerative Combinatorics*, volume 1, and the two pages preceding.)

5.3. the partition lattice

Here's another special poset.

Definition 5.21. The *partition lattice* Π_n is the poset consisting of the unordered set partitions of $[n]$, ordered by reverse refinement. (So $\sigma \preccurlyeq \tau$ if every block of σ is a block of τ.)

Here is an image of Π_4:[3]



But is Π_n actually a lattice? If $\sigma = \{\sigma_1, \ldots, \sigma_k\}$ and $\tau = \{\tau_1, \ldots, \tau_r\}$ are both set partitions of $[n]$, then their *common refinement* is the set partition whose blocks consist of the nonempty intersections $\sigma_i \cap \tau_j$. The common refinement of σ and τ is their greatest lower bound in Π_n. To define their join, let $\sim$ be the equivalence relation on $[n]$ generated by the relationships $x \sim y$ if x and y are either in the same block of σ or in the same block as τ. The partition induced by $\sim$ has both σ and τ as refinements; it is, in fact, their maximal lower bound. So Π_n is a lattice.

Π_n has a special property:

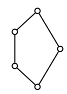
Definition 5.22. A poset $(X, \preccurlyeq)$ is *graded* if there is a function $\rho\colon X \to \mathbb{N}$ such that

1. If $x \prec y$, then $\rho(x) < \rho(y)$, and
2. If $x \lessdot y$, then $\rho(y) = \rho(x) + 1$.

The function ρ is called a *rank function* of X and $\rho(x)$ is called the *rank* of x.

[3] Image created by Tilman Piesk and placed on Wikicommons under a Creative Commons Attribution 3.0 Unported license.

The idea is that the rank defines a set of "levels" of X, and every covering relation goes between one level and the next. If X is finite, then condition (2) implies condition (1). Not every poset is graded: $(\mathbb{R}, \leq)$ is one simple example. But it's also true that not every finite poset is graded. The poset with five elements and the Hasse diagram



is not a graded poset.

But many interesting finite posets are, in fact, graded. The Boolean lattice B_n has the rank function $\rho(S) = |S|$, for example; the lattice $L_q(n)$ has the rank function $\rho(S) = \dim(S)$. The partition lattice, too, has a rank function: This one sends each set partition to n minus the number of blocks that it has. You can see the graded structure of Π_4 in the image above. The number of elements of Π_n with rank r is exactly $S(n, n-r)$. From Exercise 1.3, we know that the Stirling numbers are not symmetric, so Π_n is not a self-dual poset.

We're on a roll here, so we may as well ask the same question of this lattice: How many saturated chains does it have?

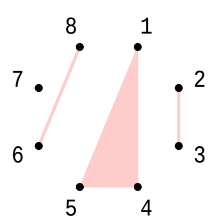
Proposition 5.23. *If $n > 1$, the partition lattice Π_n has exactly $\prod_{k=2}^{n} \binom{k}{2}$ saturated chains.*

Proof. A saturated chain is a sequence $P_0 \prec \cdots \prec P_{n-1}$ of partitions of $[n]$ where P_{k+1} is formed from P_k by merging two of its blocks into one. There are $\binom{n-k}{2}$ ways to do this; multiplying out gives the final answer. □

From here, we discuss two special subposets of Π_n.

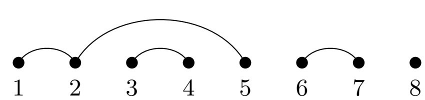
Definition 5.24. Let P_n be a regular n-gon with vertices labelled $1, 2, \ldots, n$ in clockwise order. Given a set partition $\sigma = \{\sigma_1, \ldots, \sigma_k\}$, we may associate each block σ_i to the convex hull of the corresponding vertices in P_n. We call σ *non-crossing* if none of these convex hulls intersect.

For example, the set partition $\{145, 23, 68, 7\}$ is represented as



So this particular set partition is non-crossing.

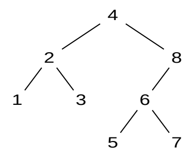
There is another way to draw partitions. Given a partition of $[n]$, we draw n dots in a line, numbered from 1 to n, and draw a line from a to b if they are consecutive numbers in a block of the partition. For example, if $\sigma = \{124, 568, 3, 7\}$ this diagram looks like



A set partition is noncrossing if and only if none of the lines in this picture cross.

Theorem 5.25. *The number of noncrossing partitions of $[n]$ is the Catalan number C_n.*

Proof sketch. Here's a simple bijection between noncrossing partitions of $[n]$ and binary trees with n vertices. To each noncrossing partition of $[n]$ associate the binary search tree that contains an edge between i and j if and only if $i < j$ are in the same block and there is no k in that block such that $i < k < j$. For example, for the noncrossing partition above, the corresponding tree is



This is a bijection; from Exercise 3.4, there are exactly C_n binary search trees, and therefore C_n noncrossing partitions. □

Exercise 5.7. Fill in the gaps in the previous proof: Why is there always a binary search tree of the described form, and why is it unique? Why is the described map a bijection?

In fact, the previous proof tells us more: Every noncrossing partition with k blocks is sent to a binary tree with $n-k$ left edges. (There is a bijection from binary trees to binary search trees that consists of removing the labels.) By Proposition 2.35, there $N(n, n-k+1)$ such trees. (The number of binary trees with k left edges is the same as the number of binary trees with k right edges.) So we get:

Proposition 5.26. *There are $N(n, n-k+1)$ noncrossing partitions with exactly k blocks.*

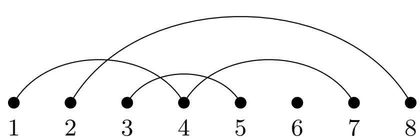
⚠ The poset of noncrossing partitions is not a sublattice of Π_n. For example: The join of the partitions $\{13, 2, 4\}$ and $\{24, 1, 3\}$, both of which are noncrossing, is the partition $\{13, 24\}$, which *is* crossing. (The common refinement of two noncrossing partitions is noncrossing, though.)

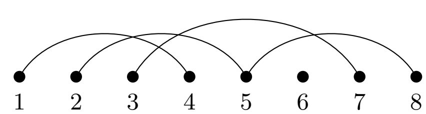
Exercise 5.8. Show that the poset of noncrossing partitions is nevertheless a lattice: Any two noncrossing partitions have a unique least upper bound. (It's simply the case that that this upper bound might not be their join in Π_n.)

The method of drawing partitions using lines gives rise to another subposet of Π_n.

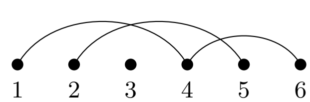
Definition 5.27. A set partition is called *nonnested* if no line in its line drawing lies entirely below another.

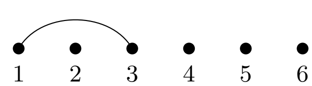
For example, the partition represented on the right is nonnested, while the one on the left is not, since the arc between 3 and 5 is contained in the arc between 2 and 8. (Singletons below a line are okay—it's only the lines that matter.)



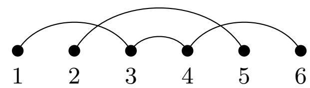


The set of nonnested partitions also forms a subposet of Π_n, but it's again not a sublattice: The join of the partitions $\{14, 2, 3\}$ and $\{23, 1, 4\}$ is nested, while neither of the original partitions is. The meet, too, doesn't work out. Take these two partitions:





Neither of them is nested, but their meet is



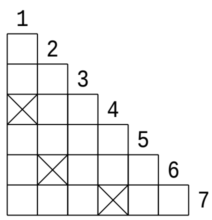
which is nested.

Exercise 5.9. Determine whether the poset of nonnested partitions forms an independent lattice; that is, does each pair of elements have a unique maximal lower bound and minimal upper bound? (Hint: Consider the previous example.)

So how many elements does this poset have?

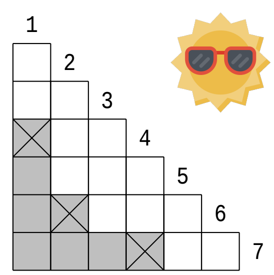
Proposition 5.28. *There are C_n nonnested partitions of $[n]$.*

Proof. We biject the set of nonnested partitions to the collection of Dyck paths.

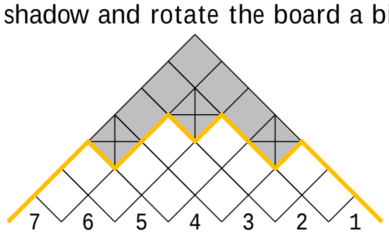


Choose some nonnested partition σ and draw its line diagram. First, draw a triangular array with $n-1$ rows and columns. Then, for each arc from i to j (with $i < j$) in the line drawing of σ, place an X in the square with the column marked i and the row marked j. For example, if the partition σ is $\{147, 26, 3\}$, we get the diagram shown on the left. The relative positions of any two X's is northwest-southeast: This is exactly equivalent to the fact that no lines are nested.

To complete the bijection, we just take this picture outside on a sunny day.[4] The sun appears in the upper right corner, as it would any elementary-school art project, and each X casts a shadow, as shown below.[5]



If you take the boundary of the shadow and rotate the board a bit, you get a Dyck path.



This process is reversible, so it is a bijection. □

Just as in Theorem 5.25, this proof actually gives more. Any partition with k blocks will have a line drawing with $n-k$ lines, in which case the triangle has $n-k$ X's, resulting in a Dyck path with $n-k+1$ peaks. So we get:

Proposition 5.29. *The number of nonnested partitions with k blocks is $N(n, n-k+1)$.*

Problem 5.10. Find a bijection between noncrossing partitions and nonnested partitions, preferably one that preserves the number of blocks.

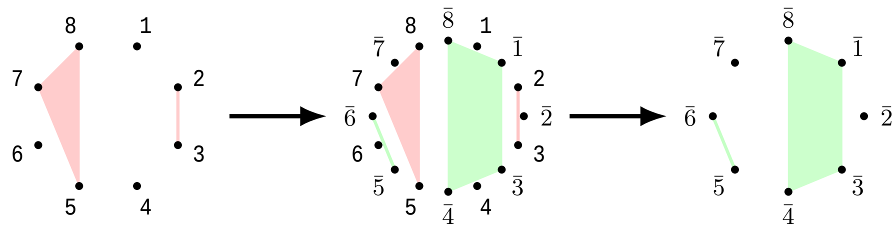
[4] Credit to Alex Postnikov for describing it this way.
[5] Icon made by <a Aranagraphics at flaticon.com.

You might wonder, based on all of this, whether these posets are actually isomorphic. They're not. If you did Exercises 5.8 and 5.9, you found that the poset of noncrossing partitions is a lattice, but the poset of nonnested partitions is not.

Noncrossing partitions, then, are a little nicer than nonnested ones. Here's one more nifty property:

Proposition 5.30. *The lattice of noncrossing partitions is self-dual.*

We'll show the bijection by example. Start with a noncrossing partition and draw its convex representation. Then draw new coordinates $\bar{1}, \ldots, \bar{n}$ with $\bar{\imath}$ between i and $i+1$, and connect as the new dots as much as possible without creating any overlaps with the first diagram. For example:



This construction is called the *Kreweras complement*.

Exercise 5.11. Convince youself that Kreweras duality is a bijection on noncrossing partitions that reverses the ordering relation.

To continue the theme, we'll count the saturated chains in this lattice. But the proof will actually use some other structures, so we'll only state it for now. For brevity, NC_n is the lattice of noncrossing partitions of $[n]$.

Theorem 5.31. *The number of saturated chains in NC_n is n^{n-2}.*

A proof appears in Section 5.5.

5.4. symmetric group lattices

Absolute order

We know that the symmetric group S_n is generated by the transpostitions $(i\ j)$. This gives us a way to measure the the complexity of a permutation, by how "mixed up" it is.

Definition 5.32. The *absolute length* or *reflection length* of a permutation $\sigma \in S_n$, denoted $\ell_T(\sigma)$, is the minimum number ℓ such that σ can be written as the composition of ℓ transopositions.

Every permutation can be written as the product of at most $n-1$ transpositions. But in fact, we can do better. Recall that every permutation can be written as a product of *cycles*.

Proposition 5.33. *$\ell_T(w)$ is equal to n minus the number of cycles in σ. (We include the cycles with one element.)*

Proof sketch. First up, any cycle $(\tau_1\ \tau_2 \cdots \tau_k)$ can be written as the product $(\tau_1\ \tau_2)(\tau_2\ \tau_3)\cdots(\tau_{k-1}\ \tau_k)$. So the length of a cycle of length k is $k-1$. If σ has r cycles, then this means that σ can be written as the product of $n-r$ transpositions.

On the other hand, multiplying a permutation τ creates at most one new cycle. So a permutation with only r cycles is the product of at least $v-r$ transpositions. $\square$

We can use this length function to define a partial order on the symmetric group.

Definition 5.34. The *absolute order* on S_n is defined by its covering relations: $\sigma \lessdot_T \tau$ if and only if there is a transposition $(i\ j)$ such that $\sigma = \tau(i\ j)$ and $\ell_T(\sigma) = \ell(T) + 1$.

Neither of these conditions implies the other; both are necessary. This is a graded poset—the absolute length provides a rank function. It is not a latice: The maximal elements of S_n under the absolute order are the so-called *long chains*: cycles of length n. There are $(n-1)!$ of these. Since this poset doesn't have a maximum element, it can't have a join operation. It does, though, have a minimum element: the identity permutation ε.

Clearly, then, this is not a self-dual poset. The structure near the top of the poset is quite different than the structure near the bottom. From a certain perspective, this is because the poset is many copies of the same lattice layered on top of each other. If we pick a single copy of that lattice, things become much simpler.

Definition 5.35. Suppose $(X, \preccurlyeq)$ is a poset and $x \preccurlyeq y$. The *interval* between x and y is $[x, y] := \{z \in X : x \preccurlyeq z \preccurlyeq y\}$.

Proposition 5.36. *If c_1 and c_2 are two long chains in S_n, then $[\varepsilon, c_1]$ is isomorphic to $[\varepsilon, c_2]$.*

Proof. Any two long chains are conjugate. (This is a special case of the fact that any two permutations with the same cycle structure are conjugate.) So there is a permutation $\sigma \in S_n$ such that $c_2 = \sigma c_1 \sigma^{-1}$. Define the map $\varphi\colon [\varepsilon, c_1] \to [\varepsilon, c_2]$ by $\varphi(\tau) = \sigma\tau\sigma^{-1}$. This preserves the absolute order and is a bijection. □

Exercise 5.12. Prove that φ actually has these properties.

For convenience, we fix the specific long chain $c := (1\ 2 \cdots n)$. Once we restrict to a specific interval, things become much nicer: $[\varepsilon, c]$ is self-dual. In fact, this follows from an isomorphism:

Theorem 5.37. *$[\varepsilon, c]$ is isomorphic to NC_n.*

Proof sketch. Here's the bijection: Take a noncrossing partition and arrange each block in increasing order. Then turn each block into a cycle by writing parentheses around it. For example: $\{135, 26, 4\} \mapsto (135)(26)(4)$. □

Exercise 5.13. Fill in the details of this proof.
(i) Why is each permutation obtained in this way a member of the interval $[\varepsilon, c]$?
(ii) Why is this map a bijection?
(iii) Why does it preserve order?

Weak Bruhat order

We can thin out the transpositions to make a much more concise generating set: The adjacent transpositions $s_i = (i\ i{+}1)$. In fact, it's possible to prove that S_n is isomorphic to the group generated by $s_1, \ldots, s_{n-1}$ modulo the relations

$$\begin{aligned} s_i s_{i+1} s_i &= s_{i+1} s_i s_{i+1} \\ s_i s_j &= s_j s_i \qquad\qquad \text{if } |i-j| \geq 2 \\ s_i^2 &= \varepsilon. \end{aligned}$$

Definition 5.38. The *adjacency length*[6] of a permutation σ, denoted $\ell(\sigma)$, is the least number ℓ such that σ can be written as the product of ℓ adjacent transpositions.

Just as before, there's an intrinsic meaning to this length.

[6] This is a nonstandard term.

Proposition 5.39. *$\ell(\sigma)$ is the number of inversions in σ.*

Exercise 5.14. Prove Proposition 5.39. (**Hint:** First prove that multiplying a permutation by an adjacent transposition changes the numer of inversions by exactly 1.)

We can copy the definition of the absolute order to get yet *another* poset.

Definition 5.40. The *Weak Bruhat order* on S_n is defined by the covering relation $\sigma \lessdot \tau$ if there is an $i \in [n-1]$ such that $\sigma = \tau s_i$ and $\ell(\sigma) = \ell(\tau) + 1$.

Exercise 5.15. Prove that this poset is a lattice.

Its minimum element is the identity permutation and its maximal element is the permutation $n, n-1, \ldots, 2, 1$.

Exercise 5.16. Prove that this lattice is self-dual.

Actually, we've already encountered this poset in some other form, as the 1-skeleton of the permutohedron. Remember that the edges of the permutohedron are correspond to permutations connected by adjacent transpositions. So we've taken this graph and given it the structure of a lattice.

Counting the saturated chains in this poset is a bit more complicated. The short answer is this:

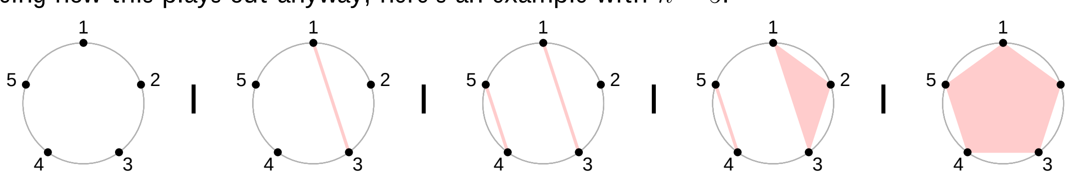
Theorem 5.41. *The number of saturated chains in S_n in the weak Bruhat order is $\binom{n}{2}! \prod_{k=1}^{n-1} (2k-1)^{-(k-1)}$.*

Every saturated chain corresponds to a "reduced decomposition" of $(n\ n{-}1 \cdots 2\ 1)$: a minimal way to write the cycle as a product of adjacent transpositions. And there's actually more to the story. These reduced decompositions are actually in bijection with a particular class of things called standard Young tableaux. But let's not get into that now.

5.5. decompositions of the long cycle

We have three different objects that are counted by n^{n-2}: labelled trees with n vertices, saturated chains in NC_n, and decompositions of the long cycle $c = (1\ 2 \cdots n)$ into a product of $n-1$ transpositions. This section provides bijections between these, which proves Theorem 5.31. We'll also allow a good amount of sidetracking.

We actually know that there is a bijection between the set of decompositions of c into transpositions and the set of saturated chains in NC_n, since the two lattices are isomorphic. It's worth seeing how this plays out anyway; here's an example with $n = 5$:

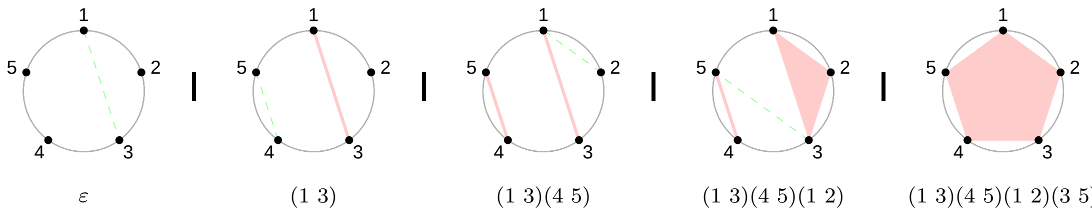


ε $\quad$ $(1\ 3)$ $\quad$ $(1\ 3)(4\ 5)$ $\quad$ $(1\ 2\ 3)(4\ 5)$ $\quad$ $(1\ 2\ 3\ 4\ 5)$

If we write it out each element of the permutation sequence as a product of transpositions, we get

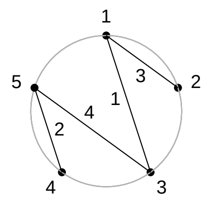
$$\varepsilon \lessdot (1\ 3) \lessdot (1\ 3)(4\ 5) \lessdot (1\ 3)(4\ 5)(1\ 2) \lessdot (1\ 3)(4\ 5)(1\ 2)(3\ 5).$$

These transpositions can actually be easily read off from the sequence of diagrams. The sequence appears again below, this time with a line from i to j drawn in when the transposition $(i\ j)$ is added.



If, in the kth step, you merge two partition blocks B_1 and B_2, then this corresponds to multiplying the permutation by the transposition $(i\ j)$, where i is the last vertex of B_1, in clockwise order, that occurs before B_2, and j is the last vertex of B_2, in clockwise order, that occurs before B_1. So that's the bijection between noncrossing partitions of $[n]$ and decompositions of c into transpositions.

Now let's go from decompositions of c to labelled trees. We can actually build on what we've already done. Here's how: Draw n evenly spaced verties in a circle. Then, given a decomposition of c, draw in the transpositions as edges and label them in the order they appear. For our running example $(1\ 3)(4\ 5)(1\ 2)(3\ 5)$, this becomes



This tree has three properties:

1. It is, in fact, a tree.
2. No two edges cross.
3. At each vertex, the labels of the edges increase in counterclockwise order.

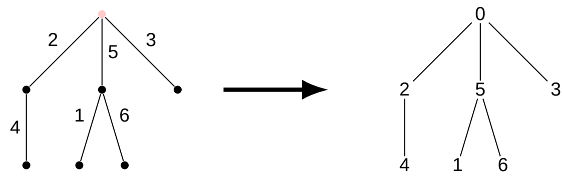
These properties are true for every labelled graph obtained in this way. For example: Any graph obtained like this from a decomposition of c has $n-1$ edges. So if it is connected, it's a tree. Otherwise, it has at least two different components, which is impossible: The corresponding permutation would have two different cycles, and c has only one. The other properties similarly follow from the construction.

Exercise 5.17. Show that any tree obtained in this way satisfies properties (2) and (3), as well.

In fact, we can go back to the construction to see that any tree with these three properties corresponds to a saturated chain in NC_n (and thus a decomposition of c). So this correspondence is a bijection.

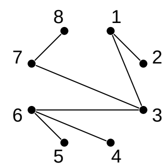
Now, each tree with n vertices and edges labelled by $1, \ldots, n-1$ can be embedded in a circle to satisfy these three conditions, and this embedding is unique up to rotation. (Take a moment to see why this is.) So these drawings in the circle are in bijection with edge-labelled trees that have a declared root note. (The bijection maps the root node to 1, which determines the rest of the mapping.)

Now we biject these edge-labelled rooted trees with vertex-labelled trees. To do this, declare the root node to have label 0; then, if e is the last edge in the unique path from the root vertex to v, set the label of v to be the label of e. For example:
This returns a graph with n vertices labelled $0, \ldots, n-1$. The process is easily reversible, so it's a bijection. And we already know (by Theorem 2.25) that there are n^{n-2} such trees, so this proves Theorem 5.31.



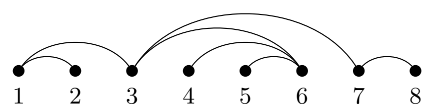
Noncrossing trees

So that's that for bijections, but now we have a new object—noncrossing trees—so let's count different them. To be more specific, a *noncrossing tree* with n vertices is a labelled tree such that, when it is drawn in a circle with labels arranged in clockwise order, no two edges cross. For example, here is one noncrossing tree:



Problem 5.18. How many noncrossing trees with n vertices are there?

We can also visualize noncrossing trees using arc diagrams; the tree above becomes



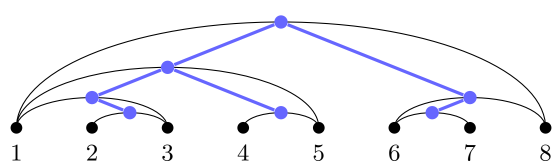
The condition that edges don't cross in the circular diagram is exactly the same as the condition that edges don't cross in the arc diagram.

Definition 5.42. A labelled tree is called *alternating* if, whenever $i_1, i_2, \ldots, i_k$ is a sequence of adjacent vertex labels, we have $i_1 > i_2 < i_3 > i_4 \cdots$, or the same with all inequalities reversed.

Exercise 5.19. Prove that a labelled tree is alternating if and only if, when each edge is directed from the smaller label to the larger label, every vertex in the tree is either a source or a sink.

Proposition 5.43. *The number of noncrossing alternating trees with n vertices is C_{n-1}.*

Proof sketch. We biject with the set of binary trees with $n-1$ vertices. In the middle of each edge in the arc diagram of the tree, place a new vertex, and connect two vertices if one is directly below the other. For example:



□

Exercise 5.20. Show that any noncrossing alternating tree contains an edge betweeen 1 and n. Then use this to fill out the proof of Proposition 5.43.

Problem 5.21. A labelled tree is called *non-nesting* if in its arc diagram, there is not a pair of edges ac and bd with $a < b < c < d$. Show that the number of non-nesting alternating trees is also C_{n-1}.

Hurwitz problems

We can restate the result on decompositions of the long cycle like this:

Proposition 5.44. *The number of $(n-1)$-tuples $(t_1, \ldots, t_{n-1})$ of transpositions in S_n such that $t_1 \cdots t_{n-1}$ is an n-cycle is $(n-1)!n^{n-2}$.*

This problem has been *grandly* generalized to statements about decompositions of "permutations of this type" into "permutations of that type." Counting problems like this are called *Hurwitz problems*. The most general result is the ESLV formula, named after its authors Torsten Ekedahl, Sergei Lando, Michael Shapiro, and Alek Vainshtein; even stating it would be much too difficult for these notes. Instead, we'll state result on decompositions into transpositions that's still quite general.

Definition 5.45. Let σ be a permutation, and write it as a product of disjoint cycles such that the length of the cycles is nonincreasing. The ordered list $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell$ of the lengths of the cycles is called the *cyclic type* of σ.

Theorem 5.46. *Let $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell > 0)$ be a partition of n, and let m_i be the number of parts of λ that are equal to i. We call a k-tuple $(t_1, \ldots, t_k)$ of transpositions in S_n nifty[7] if*

- *$t_1 \cdots t_k$ has cyclic type λ,*
- *$t_1, \ldots, t_k$ generate S_n, and*
- *$k = n + \ell - 2$.*

The number of nifty k-tuples is

$$n^{\ell-3} k! n! \prod_{i \geq 1} \frac{1}{m_i} \left(\frac{i^i}{i!} \right)^{m_i}.$$

There's a lot to unpack here. First up: The number $n + \ell - 2$ is the minimum number of transpositions needed to satisfy the first two conditions. This is not obvious; if you want, you can try to prove it. Next up, why require that the t_i generate S_n? If we didn't require that, then the corresponding theorem would have $k = n - \ell$ (the minimum number required to produce a permutation with cyclic type λ) and would follow from the theorem on the long cycle—you just treat each cycle independently. Finally, a bit of fun:

Exercise 5.22. Given a set of transpositions in S_n, we can form a labelled graph on n vertices that contains an edge between i and j if the set of transpositions contains the transposition $(i\ j)$. Show that this tree is connected if and only if the set of transpositions generates S_n.

Kreweras complement of trees and cycle decompositions

Remember that the Kreweras complement of a noncrossing partition is an order-preserving map from NC_n to its dual. If π is a noncrossing partition, we'll use π^K to denote its Kreweras complement. You can check that if

$$\hat{0} = \pi_0 \lessdot \pi_1 \lessdot \cdots \lessdot \pi_n = \hat{1}$$

is a saturated chain in NC_n, then

$$\hat{1} = \pi_0^K \gtrdot \pi_1^K \gtrdot \cdots \gtrdot \pi_n^K = \hat{0}.$$

So the Kreweras complement induces a bijection (in fact, an involution) on the set of saturated chains in NC_n. We have bijections from NC_n to the set of decompositions of the long cycle and the set of labelled trees; what sort of map does the Kreweras dual induce in these objects?

Problem 5.23. Ponder this.

[7] This is nonstandard.

5.6. möbius inversion

In this section, we introduce a new algebraic gadget for posets. In some ways, it's a very far-reaching generalization of the principle of inclusion-exclusion, so let's recall that first.

Proposition 5.47 (Principle of inclusion-exclusion). *If $A_1, \ldots, A_n$ are finite sets, then*

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{\emptyset \neq J \subseteq [n]} (-1)^{|J|-1} \left|\bigcap_{j \in J} A_j\right|.$$

If $n = 2$, this can be written in the more familiar form

$$|A \cup B| = |A| + |B| - |A \cap B|;$$

and $n = 3$ is

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

We'll circle back to this, so put it to simmer aromatically on the back burner while we take a turn toward the algebraic.

Definition 5.48. If $(P, \preceq)$ is a poset and $x, y \in P$ with $x \preceq y$, the *interval* between x and y is the set

$$[x, y] := \{z \in P : x \preceq z \preceq y\}.$$

If this set is finite for every $x, y \in P$, then we call P *locally finite*.

Definition 5.49. Suppose $(P, \preceq)$ is a locally finite poset and k is a field, and let $\mathrm{int}(P)$ denote the set of intervals in P. The *incidence algebra* of P on k, denoted $I_k(P)$ or $I(P)$ if the field understood, contains as elements the set of functions $\mathrm{int}(P) \to k$; if $f \in I(P)$, we write $f(x, y)$ for $f([x, y])$. Given two functions $f, g \in I(P)$, their sum is the function

$$(f + g)(x, y) = f(x, y) + g(x, y)$$

and their product is

$$(f * g)(x, y) = \sum_{x \preceq z \preceq y} f(x, z)g(z, y).$$

The definition of the product of two functions is odd, but the idea is that we want to tie the structure of the poset into the algebraic structure of $I(P)$; pointwise multiplication wouldn't do that. (Also, because P is locally finite, the sum on the right hand side is finite, so multiplication is well-defined.)

There are two other ways of thinking about the incidence algebra.[8] We can think of $I(P)$ as instead the vector space $k^{\mathrm{int}(P)}$; that is, the vector space with intervals $[x, y]$ as a basis. Multiplication of these basis elements is defined as

$$[x, s] * [t, y] = \begin{cases} [x, y] & \text{if } s = t \\ 0 & \text{otherwise.} \end{cases}$$

You can check that these descriptions are the same.

There's even one more, if P is finite: List the elements in some order $x_1, \ldots, x_n$ so that $i < j$ whenever $x_i < x_j$. (This is called a *linear extension* of P.) Then we can make any function

[8] Incidentally,[9] $I(P)$ *is* an algebra in the commutative algebra sense of the word. We'll see soon that $I(P)$ has an identity element δ, so the map $k \to I(P)$ given by $a \mapsto a\delta$ is a ring homomorphism that makes $I(P)$ into a k-algebra. If this makes no sense to you, then you can happily ignore it.

[9] Ha!

$f(x,y) \in I(P)$ can be written as a matrix $\big(f(x,y)\big)_{x,y\in P}$, where $f(x,y) = 0$ if $x \not\preceq y$. Each of these matrices is upper-triangular, so in this way we can embed $I(P)$ into the matrix ring $k^{n\times n}$. For example, if P is the Boolean lattice B_2, we can list its elements as $\emptyset, \{1\}, \{2\}, \{1,2\}$, and the elements of $I(P)$ correspond to matrices of the form

$$\begin{pmatrix} * & * & * & * \\ 0 & * & 0 & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Moreover, if A and B are the matrices corresponding to $f, g \in I(P)$, respectively, then AB is the matrix corresponding to $f * g$.

So we have three different ways of looking at this incidence algebra, and each of them is useful at different times. The matrix perspective, for example, makes it patently clear that $I(P)$ has an identity element: the function corresponding to the identity matrix.

Definition 5.50. The delta function $\delta \in I(P)$ is defined

$$\delta(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 5.51. *The delta function is the identity element of $I(P)$.*

Now that we have an identity element, we can ask which elements of $I(P)$ have inverses. Again, the matrix perspective helps: An upper triangular matrix is invertible if and only if every diagonal entry is nonzero. This means that if $f \in I(P)$ is invertible, then necessarily $f(x,x) \neq 0$ for every $x \in P$. But this is not obviously sufficient: Matrices in $I(P)$ have entries above the diagonal which must also be zero (as in the example with B_2 above), and it's not clear that matrix inversion preserves this property.

In fact it does, when these 0-positions are determined by a poset, at least; this is the content of the next theorem.

Theorem 5.52. *For each element $f \in I(P)$, the following are equivalent:*

1. *f has a left inverse.*
2. *f has a right inverse.*
3. *f has a two-sided inverse.*
4. *$f(x,x) \neq 0$ for every $x \in P$.*

Proof. We have already seen that each of the first three conditions implies the fourth. Now we show that the fourth implies the rest. So: If $f(x,x) \neq 0$ for every $x \in P$, we can define the function $g \in I(P)$ by

$$g(x,y) = -f(x,x)^{-1} \sum_{x \prec z \preceq y} f(x,z)g(z,y)$$

if $x \prec y$ and $g(x,x) = f(x,x)^{-1}$. (If $x \succ y$, then of course $h(x,y) = 0$.)[10] By rearranging, we get $(f * g)(x,y) = 0$ whenever $x \neq y$, and $(f * g)(x,x) = 1$ for every $x \in P$. In other words, $f * g = \delta$, so f has a right inverse. A similar argument shows that f has a right inverse h. Since $h = h * f * g = g$, we see that f has a two-sided inverse. □

Now we'll zero in on one particular element in the incidence algebra.

[10] This particular definition might seem mysterious, but it's not meant to be: It's exactly what you get by assuming that f has a right inverse, writing out the formula $f * g$, and solving for g.

Definition 5.53. The *zeta function* of $I(P)$ is

$$\zeta_P(x,y) = \begin{cases} 1 & \text{if } x \preccurlyeq y \\ 0 & \text{otherwise.} \end{cases}$$

We write $\zeta(x,y)$ when the poset is understood.

Exercise 5.24. Show that $\zeta^2(x,y)$ is the number of elements in the interval $[x,y]$.

Suppose we have a function $f\colon P \to k$. We can turn this into an element of the incidence algebra by setting $\bar{f}(x,y) = f(x)$ if $x \preccurlyeq y$ and 0 otherwise. In this case,

$$(\zeta * \bar{f})(x,y) = \sum_{x \preccurlyeq z \preccurlyeq y} \bar{f}(z,y) = \sum_{x \preccurlyeq z \preccurlyeq y} \bar{f}(z).$$

So ζ is acting as a kind of discrete integral; if we were fanciful, we could write

$$(\zeta * \bar{f})(x,y) = \int_{[x,y]} f(z)\, dz.$$

Of course, we are fanciful, which is why we wrote it.

Anyway, back to the zeta function. Since $\zeta(x,x) = 1$ for every $x \in P$, it's invertible. Its inverse is called the *Möbius function* of P; it's denoted μ_P or simply μ. This function leads us to a very important tool. (It's the title of this section, after all.) Let's call a poset *downward finite* if $\{y \in P : y \preccurlyeq x\}$ is finite for every $x \in P$.[11]

Theorem 5.54 (Möbius inversion). *Suppose that P is a downward finite poset.*[12] *If $f, g\colon P \to K$, then*

$$g(y) = \sum_{x \preccurlyeq y} f(x) \quad \text{if and only if} \quad f(y) = \sum_{x \preccurlyeq y} g(x)\mu(x,y).$$

Proof. Form a new poset $\hat{P}$ by adding a minimum element $\hat{0}$ to P. We define two functions $\hat{f}, \hat{g} \in I(\hat{P})$ by

$$\hat{f}(x,y) = \left\{ \begin{array}{ll} f(y) & \text{if } x = \hat{0} \\ 0 & \text{otherwise} \end{array} \right\} \quad \text{and} \quad \hat{g}(x,y) = \left\{ \begin{array}{ll} g(y) & \text{if } x = \hat{0} \\ 0 & \text{otherwise} \end{array} \right\}.$$

Then $g(y) = \sum_{x \preccurlyeq y} f(x)$ if and only if $\hat{g} = \hat{f} * \zeta$; and $f(y) = \sum_{x \preccurlyeq y} g(x)\mu(x,y)$ if and only if $\hat{f} = \hat{g} * \mu$. (Verify this!) But we know that $\hat{g} = \hat{f} * \zeta$ if and only if $\hat{f} = \hat{g} * \mu$, which finishes the proof. □

Sometimes it's useful to reverse the inequality symbols. Then we get this:

Theorem 5.55 (Möbius inversion, dual). *Suppose that P is an upward finite poset.*[13] *If $f, g\colon P \to K$, then*

$$g(y) = \sum_{x \succcurlyeq y} f(x) \quad \text{if and only if} \quad f(y) = \sum_{x \succcurlyeq y} g(x)\mu(x,y).$$

Proof. This statement is true if and only if Theorem 5.54 is true in the dual poset of P. □

In *Enumerative Combinatorics*, Stanley provides a more algebraic proof of Theorem 5.54.

[11] The standard way of describing this is that "every principal order ideal of P is finite."
[12] This is just so that the sums make sense.
[13] That is, $\{y \in P : y \succcurlyeq x\}$ is finite for every $x \in P$.

It turns out that we can continue the calculus analogy that was gestured at before. Given $a, b \in P$ with $a \preccurlyeq b$ and a function $f\colon P \to k$, let's formally define a "discrete integral" as

$$\int_a^b f(z)\,dz = \sum_{a \preccurlyeq z \preccurlyeq b} f(z),$$

and a "discrete derivative" as

$$(D_a f)(y) = \sum_{a \preccurlyeq z \preccurlyeq y} f(z)\mu(z, y).$$

Then, noting that a is the minimum element in the interval $[a, b]$ and applying Theorem 5.54 to the subposet $[a, b] \subseteq P$, we get

$$f(y) = D_a \int_a^y f(z)\,dz \qquad \text{and} \qquad g(y) = \int_a^y (D_a g)(z)\,dz.$$

These are some kind of discrete analogue of the fundamental theorems of calculus. You could go farther, developing a product and quotient rule, integration by parts, and so on, but (as far as I can tell) no one has done that, except in the very special case of the poset $(\mathbb{Z}, \leq)$, which we'll see more of in a bit. This is not to be confused with the other idea of calculus on posets.

Let's get back to Möbius inversion writ large. To apply Theorem 5.54 in concrete scenarios, and to see why inclusion-exclusion is actually a special case, we'll have to be able to calculate values of the Möbius function.

5.7. the möbius function

Already in Theorem 5.52, we can see the seeds for a formula for a recurrence relation for the Möbius function.

Proposition 5.56. *The values of the Möbius function are given by the recurrence $\mu(x, x) = 1$ for all $x \in P$ and*

$$\mu(x, y) = -\sum_{x \prec z \preccurlyeq y} \mu(z, y) = -\sum_{x \preccurlyeq z \prec y} \mu(x, z)$$

if $x \prec y$.

Proof. The first sum comes from expanding out the expression for $\zeta * \mu = \delta$; the second comes from expanding out $\mu * \zeta = \delta$. □

This is enough to calculate the Möbius function of some very simple posets.

Example 5.57. The Möbius function on $P = (\mathbb{Z}, \leq)$ is

$$\mu(x, y) = \begin{cases} 1 & \text{if } x = y \\ -1 & \text{if } y = x + 1 \\ 0 & \text{otherwise.} \end{cases}$$

To prove this, just calculate using Proposition 5.56. The same formula hold for any finite chain.

Let's go back to the discrete calculus we defined before for the special case $(\mathbb{Z}, \leq)$. Since we know the Möbius function explicitly, we find that

$$(D_a f)(y) = \sum_{a \preccurlyeq z \preccurlyeq y} f(z)\mu(z, y) = f(y) - f(y - 1)$$

so long as $a < y$. On $\mathbb{Z}$, then, we can define the *discrete derivative* of a function $f\colon \mathbb{Z} \to k$ by $\Delta f(n) = f(n) - f(n - 1)$, which no longer depends on any base point. (This should hopefully appear like a natural approximation of a derivative on the discrete set $\mathbb{Z}$.) In this case, the analogy with calculus has been carried even further; see these notes for a primer.

To calculate more complex lattices, we'll need more tools. Here's a simple one.

Definition 5.58. If $(P, \preccurlyeq_P)$ and $(Q, \preccurlyeq_Q)$ are posets, their *product poset* $(P \times Q, \preccurlyeq_{P\times Q})$ has the partial order defined by $(a, b) \preccurlyeq_{P\times Q} (c, d)$ if and only if $a \preccurlyeq_P c$ and $b \preccurlyeq_Q d$.

Proposition 5.59. *If P and Q are locally finite posets, then $\mu_{P\times Q}\big((a, b), (c, d)\big) = \mu_P(a, c) \times \mu_Q(b, d)$.*

Proof. Note that $I(P \times Q)$, considered as a k-algebra, is generated by the elements $\big[(a, b), (c, d)\big]$ with $a \preccurlyeq_P c$ and $b \preccurlyeq_Q d$, that is, with $[a, c] \in I(P)$ and $[b, d] \in I(Q)$. This shows that $I(P \times Q) \cong I(P) \otimes_k I(Q)$. It's easy to check that this maps sends $\zeta_{P\times Q}$ to $\zeta_P \otimes \zeta_Q$; taking inverses, we get that $\mu_{P\times Q} \mapsto \mu_P \otimes \mu_Q$. Evaluating at the interval $\big[(a, b), (c, d)\big]$ in $P \times Q$ finishes the proof. □

If this proof is a little too algebra-heavy for you, Stanley also provides a more straightforward proof by calculation.

Corollary 5.60. *The Möbius function of the Boolean lattice B_n is given by*

$$\mu(S, T) = (-1)^{|T \setminus S|}$$

if $T \supseteq S$.

Proof. Let P_2 denote the lattice with 2 elements $\hat{0} < \hat{1}$. Then B_n is isomorphic to $(P_2)^n$ via the isomorphism that sends a set S to the vector $v^S \in \{0, 1\}^n$ with $v^S_i = 1$ if and only if $i \in S$. The product formula tells us that

$$\mu(S, T) = (-1)^{\|v^T - v^S\|_1} = (-1)^{|T \setminus S|}.$$ □

And now we come to the promised corollary: inclusion-exclusion.

Proof of the principle of inclusion-exclusion. Let $A_1, \ldots, A_n$ be finite sets, and set $X = \bigcup_{i=1}^n A_i$. We define two functions $f, g \colon B_n \to \mathbb{R}$ by

$$f(J) = \Big|\bigcap_{j\in J} A_j\Big| \qquad \text{and} \qquad g(J) = \Big|\Big(\bigcap_{j\in J} A_j\Big) \cap \Big(\bigcap_{j\notin J} (X \setminus A_j)\Big)\Big|.$$

You can check that for any $I \subseteq [n]$, we have $f(I) = \sum_{J\supseteq I} g(I)$. Theorem 5.55 together with Corollary 5.60 implies that $g(I) = \sum_{J\supseteq I} (-1)^{|J\setminus I|} f(J)$ for every $I \subseteq [n]$. Evaluating at $I = \emptyset$, we get

$$0 = g(\emptyset) = \sum_{J\subseteq [n]} (-1)^{|J|} f(J).$$

Solving for $f(\emptyset)$ finishes the proof. □

Number theory has its own thing called Möbius inversion.

Proposition 5.61 (Möbius inversion, number theory). *Define the function $\mu \colon \mathbb{N} \to \{0, \pm 1\}$ by the following rule:*

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is square-free and the product of } k \text{ distinct primes} \\ 0 & \text{if } n \text{ is divisible by a square.} \end{cases}$$

For any two functions $f, g \colon \mathbb{N} \to k$, we have

$$f(n) = \sum_{d|n} g(d) \qquad \text{if and only if} \qquad g(n) = \sum_{d|n} f(d)\mu(n/d).$$

This, too, is a special case of our poset inversion. In particular: Let DL denote the poset of the natural numbers ordered by divisibility (so $m \preccurlyeq_D n$ if $m \mid n$; we'll just write the latter).

Exercise 5.25. Show that DL is isomorphic to the direct sum of $(\mathbb{N}_0, \leq)$ countably many times. (**Hint:** Let e_k denote the element in $\bigoplus_{i \in \mathbb{N}}(\mathbb{N}_0, \leq)$ that has 1 in its kth coordinate and 0 everywhere else. If p_k denotes the kth prime number, the map that sends $m = p_{i_1}^{a_1} \cdots p_{i_r}^{a_r}$ to $\sum_{j=1}^{r} a_r e_{i_j}$ is an order-preserving isomorphism.)

Exercise 5.26. Show that D is a lattice; thus the initialism DL. (What are its meet and join operations?)

Using this correspondence, Proposition 5.59 tells us that

$$\mu_{DL}(m,n) = \begin{cases} (-1)^k & \text{if } n/m \text{ is square-free and is divisible by } k \text{ distinct primes} \\ 0 & \text{if } m/n \text{ is divisible by a square.} \end{cases}$$

So $\mu(m/n) = \mu_{DL}(m,n)$, and Proposition 5.61 is nothing more than Möbius inversion applied to the divisor lattice.

In the divisor lattice, the discrete derivative is largely independent of the base point. The *radical* of a positive integer n, denoted $\operatorname{rad}(n)$, is the product of the distinct primes that divide it. (So, for example, $\operatorname{rad}(180) = 2 \cdot 3 \cdot 5 = 30$.) Let $\omega(n)$ denote the number of distinct prime factors of n. If $a \leq \operatorname{rad}(n)$, then

$$\big(D_a f\big)(n) = \sum_{a|d|n} f(d)\mu(n/d) = \sum_{q|\operatorname{rad}(n)} (-1)^{\omega(q)} f(n/q),$$

since if n/d is divisible by a square, then $\mu(n/d) = 0$. In particular, $D_a = D_1$ for every $a \in \mathbb{N}$. So the divisor lattice has a single discrete derivative, which we can denote simply by D, and a discrete integral

$$\int_a^b f(k)\,dk = \sum_{a|k|b} f(k).$$

Of course, this can be rephrased in terms of the isomorphic lattice $\mathbf{N} = \bigoplus_{i \in \mathbb{N}}(\mathbb{N}, \leq)$. Here, $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for every $i \in \mathbb{N}$. The discrete integrals and derivatives are

$$\int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x})\,d\mathbf{x} = \sum_{\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}} f(\mathbf{x}) \qquad \text{and} \qquad \big(Df\big)(\mathbf{x}) = \sum_{\mathbf{0} \leq \mathbf{z} \leq \mathbf{1_x}} (-1)^{|\mathbf{z}|} f(\mathbf{x} - \mathbf{z}),$$

where $\mathbf{1_x}$ is the vector with 1 in the ith coordinate if $x_i \neq 0$ and 0 in the ith coordinate otherwise; $|\mathbf{z}|$ is the number of nonzero coordinates of $\mathbf{z}$.[14]

We know that discrete analogues of the fundamental theorems of calculus exist in this lattice. Can this analogy be extended further? Probably, but I don't know.

Problem 5.27. Think about this.

For fun, this Wikipedia page has yet another discrete variant of calculus.

6. hyperplane arrangements

A hyperplane arrangement is simply a collection of hyperplanes. More precisely, a hyperplane arrangement over a field F is a collection of affine $(n-1)$-dimensional subspaces of F^n.

[14] We can also look at a modified derivative in the poset $\bigoplus_{i=1}^{n}(\mathbb{Z}, \leq)$. The discrete integral is the same as the integral in DL, but we define the derivative as

$$(Df)(\mathbf{x}) = \sum_{\mathbf{0} \leq \mathbf{z} \leq \mathbf{1}} (-1)^{|\mathbf{z}|} f(\mathbf{x} - \mathbf{z}),$$

where $\mathbf{1}$ is the vector with every coordinate equal to 1. If $n = 1$, this is the discrete calculus on $\mathbb{Z}$ that we mentioned before, and if $n > 1$, this gives some sort of "multivariable discrete calculus."

6.1. graphical arrangements

From any hyperplane arrangement, we can extract a poset.

Definition 6.1. If $\mathcal{A}$ is a collection of hyperplanes in F^n for some field F, its *intersection poset* $L_{\mathcal{A}}$ consists of the nonempty intersections of hyperplanes in $\mathcal{A}$ ordered by reverse inclusion.

Example 6.2. Suppose that $\mathcal{A} = \{H_1, H_2\}$ consists of two non-parallel hyperplanes in $\mathbb{R}^n$. The intersection lattice $L_{\mathcal{A}}$ looks like this:

$$\begin{tikzcd}[column sep=small] & H_1 \cap H_2 \arrow[dl, dash] \arrow[dr, dash] & \\ H_1 \arrow[dr, dash] & & H_2 \arrow[dl, dash] \\ & \hat{0} = \mathbb{R}^n & \end{tikzcd}$$

An intersection lattice is not, in general, a lattice. If it contains two parallel hyperplanes, for example, then these two hyperplanes have no common upper bound in $L_{\mathcal{A}}$. A hyperplane arrangement is called *central* if every hyperplane in the arrangment contains the origin.

Exercise 6.1. Show that the intersection poset of any central hyperplane arrangement is a lattice.

Central hyperplane arrangements are common, making intersection lattices common—the L in $L_{\mathcal{A}}$ stands for "lattice."

Graphs parametrize a specific class of hyperplane arrangements.

Definition 6.3. Given a graph G and a field F, the *graphical arrangement* $\mathcal{A}_{G,F}$ is the hyperplane arrangement in $F^{V(G)}$ with a hyperplane H_e for each edge $uv \in E(G)$ defined by

$$H_e = \{\mathbf{x} \in F^{V(G)} : x_u = x_v\}.$$

Every graphical arrangement is central, so its intersection poset is a lattice.

Example 6.4. If G is the path graph on three vertices •–•–•, then the intersection lattice $L_{\mathcal{A}_{G,\mathbb{R}}}$ is the one that appears in Example 6.2.

It turns out that $L_{\mathcal{A}_{G,F}}$ is the same regardless of the field F. This is not true for arbitrary arrangements: If $H_1 = \{x : x_1 = 0\}$ and $H_2 = \{x : x_1 - 2x_2 = 0\}$, then interpreted as hyperplanes in $\mathbb{R}^n$, the intersection lattice looks like a diamond. But over $\mathbb{F}_2$, these hyperplanes are the same, and their intersection lattice looks like this:

$$\begin{tikzcd} H_1 = H_2 \arrow[d, dash] \\ \hat{0} = \mathbb{R}^n \end{tikzcd}$$

Essentially, this is because the hyperplanes defining $\mathcal{A}_{G,F}$ only depend on equality. Actually:

Proposition 6.5. *$L_{\mathcal{A}_{G,F}}$ is isomorphic to the sublattice of Π_n consisting of all set partitions $\{B_1, \ldots, B_k\}$ such that $G|_{B_i}$ is connected for every $i \in [k]$.*

Exercise 6.2. Prove Proposition 6.5.

This sublattice of Π_n is completely independent of the field F. Because of all of this, and to avoid triple subscripts (or even quadruple subscripts, as appears in $L_{\mathcal{A}_{K_n,\mathbb{R}}}$), we abbreviate $L_{\mathcal{A}_{G,F}}$ by simply L_G.

6.2. möbius functions, chromatic polynomials, and hyperplane arrangements, oh my!

We'll start with something completely different.

Definition 6.6. A *q-coloring* of a graph G is a function $V(G) \to [q]$. A coloring φ is called *proper* if $\varphi(u) \neq \varphi(v)$ whenever $uv \in E(G)$. The number of proper q-colorings of G is denoted $\chi_G(q)$.

The *chromatic number* of G is the least q such that $\chi_G(q) > 0$, for example. It seems like χ_G could be any sort of function, but surprisingly, it has a fairly simple structure.

Theorem 6.7. *For every graph G, there is an integer polynomial f_G such that $\chi_G(q) = f_G(q)$ for every $q \in \mathbb{N}$.*

For this reason, χ_G is called the *chromatic polynomial* of G.

Proof of Theorem 6.7. We use the so-called *deletion-contraction method*. Given an edge $e \in E(G)$, let $G \setminus e$ denote the graph obtained by deleting the edge e, and let G/e denote the graph obtained by *contracting* e, that is, by deleting e and joining its two endpoints into a single vertex.

For every $q \in \mathbb{N}$, we have

$$\chi_{G\setminus e}(q) = \chi_G(q) + \chi_{G/e}(q).$$

To see this, consider the q-colorings of $G \setminus e$. If the endpoints of e are have the same color, then this is a valid coloring of G/e. If the endpoints of e have different colors, then this is a valid coloring of G. (And the reverse inclusions are true, as well.) This shows the equality. Rearranging, we get

$$\chi_G(q) = \chi_{G\setminus e}(q) - \chi_{G/e}(q),$$

where $G \setminus e$ and G/e both have one fewer edge than G. We can finish the proof by inducting on the number of edges, so long as we know that χ_G is a polynomial whenever G has no edges. But this is clear: If G has no edges, every q-coloring is proper, so $\chi_G(q) = q^{|V(G)|}$. □

Corollary 6.8. *χ_G is a monic polynomial whose constant term is 0.*

Proof. The chromatic polynomial for every edgeless graph is divisible by q, which implies that χ_G is divisible by q for any graph G. So its constant term is 0.

Suppose that G has n vertices. Any q-coloring $V(G) \to [q]$ that is injective is proper, so $\chi_G(q) \geq q^{\underline{n}} = q(q-1)\cdots(q-n+1)$. On the other hand, $\chi_G(q) \leq q^n$, so

$$\lim_{q\to\infty} \frac{\chi_G(q)}{q^n} = 1,$$

which shows that the leading term of $\chi_G(q)$ is q^n. □

A coloring of the complete graph K_n is proper precisely when it is injective. Therefore $\chi_{K_n}(q) = q^{\underline{n}} = q(q-1)\cdots(q-n+1)$. If we write

$$q^{\underline{n}} = \sum_{k=1}^{n} s(n,k) q^k,$$

the coefficients $s(n,k)$ are called the *Stirling numbers of the first kind*.

Exercise 6.3. Show that $s(n,k) = (-1)^{n-1}\big|\{\sigma \in S_n : \sigma \text{ has } k \text{ cycles}\}\big|$.

This means that $s(n,1) = (-1)^{n-1}(n-1)!$, so $[q]\chi_{K_n}(q) = (-1)^{n-1}(n-1)!$. So far, so good.

What does all of this have to do with the Möbius function? Well,

Proposition 6.9. $\mu_{\Pi_n}(\hat{0},\hat{1}) = (-1)^{n-1}(n-1)!$.

Recall that Π_n is the partition lattice of $[n]$.

Proof. As a preliminary, note that if you pick any element $\pi \in \Pi_n$ with k blocks, the interval $[\pi, \hat{1}]$ is isomorphic to Π_k, since every element in the interval is a set partition of the blocks of π. (In fact, every interval $[\sigma, \pi]$ in Π_n is isomorphic to a product of partition lattices: If π has blocks $L_1, \dots, L_r$ and there are k_i blocks of σ inside L_i, then $[\sigma, \pi] \cong \Pi_{k_1} \times \cdots \times \Pi_{k_r}$.)

The proof now proceeds by induction. For $n = 1$, we have $\hat{1} = \hat{0}$, so $\mu(\hat{0}.\hat{1}) = 1$. Now suppose the proposition is true for all $k < n$ with $n > 1$. From the Möbius recurrence formula we have

$$\mu_{\Pi_n}(\hat{0}, \hat{1}) = -\sum_{\pi \neq 0} \mu_{\Pi_n}(\pi, \hat{1}).$$

If π has k blocks, then $[\pi, \hat{1}] \simeq \Pi_k$, so $\mu_{\Pi_n}(\pi, \hat{1}) = \mu_{\Pi_k}(\hat{0}, \hat{1}) = (-1)^{k-1}(k-1)!$. There are $S(n,k)$ elements of Π_n that have exactly k blocks, so

$$\mu_{\Pi_n}(\hat{0}, \hat{1}) = -\sum_{k=1}^{n-1} (-1)^{k-1}(k-1)!\, S(n,k).$$

To prove the theorem, we need to show that

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

Ignoring the sign for the moment, the terms on the left side count the number of cyclically ordered set partitions of $[n]$. To show that the sum is 0, we need to show that there are exactly as many even cyclically ordered set partitions as there are odd. Define a map φ on the set of cyclic set partitions such that

- if $\{n\}$ is a block in the cyclic partition π, then $\varphi(\pi)$ is the cyclic partition formed by joining $\{n\}$ with the block that immediately precedes.
- if $\{n\}$ is not a block in π, then $\varphi(\pi)$ is obtained by deleting n from the block it is in and inserting the new block $\{n\}$ between the block that formerly contained n and its successor in the cyclic ordering.

Then φ is an involution on cyclic partitions and interchanges even and odd blocks.

Alternatively, taking $t = -1$ in Corollary 4.23 shows immediately that the sum vanishes. Either way, the sum vanishes, which completes the proof. □

Remember that the intersection lattice of the graphical arrangement of K_n is isomorphic to Π_n. This suggests that there's a connection between Möbius functions and the chromatic polynomial. Indeed there is.

Proposition 6.10. *For any graph G with n vertices and any prime power q, we have*

$$\chi_G(q) = \Big| \mathbb{F}_q^n \setminus \bigcup_{H \in \mathcal{A}_G} H \Big|.$$

Proof. Just unravel the definitions. Every point in $\mathbb{F}_q^n$ represents a q-coloring of G; this coloring is improper precisely when it lies inside one of the hyperplanes in $\mathcal{A}_G$. □

And here it is:

Theorem 6.11. *For every graph G and every positive integer q, we have*

$$\chi_G(q) = \sum_{T \in L_G} \mu_{L_G}(\hat{0}, T) q^{\dim T},$$

where $\dim T$ denotes the dimension of the space T.

Proof. First assume that q is a power of a prime. We use Möbius inversion on the lattice L_G, where we consider the graphical arrangement $\mathcal{A}_G$ over the finite field F_q. Define $f, g \colon L_G \to \mathsf{N}$ by $f(S) = |S|$ and $g(S) = \left|S \setminus \bigcup_{T \subsetneq S} T\right|$. Then

$$f(S) = \sum_{T \subseteq S} g(T),$$

so by Möbius inversion

$$g(S) = \sum_{T \subseteq S} \mu_{L_G}(S, T) f(T).$$

From Proposition 6.10, we have

$$\chi_G(q) = g(\hat{0}) = \sum_{T \in L_G} \mu_{L_G}(\hat{0}, T) q^{\dim T}.$$

This is an equality of polynomial values that holds for infinitely many inputs, so it is an equality on the level of polynomials, as well. Thus it holds for any value of q. □

6.3. characteristic polynomial of hyperplane arrangements

Using Theorem 6.11 as inspiration, we can define a polynomial for any hyperplane arrangement.

Definition 6.12. If $\mathcal{A}$ is a collection of hyperplanes (not necessarily central) in F^n, its *characteristic polynomial* is

$$\chi_{A}(t) = \sum_{S \in L_{\mathcal{A}}} \mu_{L_{\mathcal{A}}}(\hat{0}, S) t^{\dim(S)}.$$

Alright. It's a polynomial. Conveniently, "chromatic" and "characteristic" both begin with *ch*, so χ works for both. And at least one interpretation remains the same.

Proposition 6.13. *If q is a power of a prime and $\mathcal{A}$ is a hyperplane arrangement in F_q^n, then*

$$\chi_{A}(q) = \left|\mathsf{F}_q^n \setminus \bigcup_{H \in A} H\right|.$$

Exercise 6.4. Prove Proposition 6.13 by following the proof of Theorem 6.11.

Typically, we think of hyperplane arrangements in R^n. For these arrangements, it's not very useful to count the points that aren't in the hyperplanes (there are, well, *uncountably* many). It's possible, however, that we can obtain information about a real hyperplane arrangement by passing to a finite field.

Every hyperplane $H \subseteq \mathsf{R}^n$ has the form

$$H = \{x \in \mathsf{R}^n : a_1 x_1 + \cdots a_n x_n = c\}$$

for some $a_1, \ldots, a_n, c \in \mathsf{R}$; H is called *rational* if each of these numbers is. If H is rational, then by multiplying by a common denominator or dividing by the least common multiple, we can ensure that every coefficient in this equation is an integer and that $\gcd(a_1, \ldots, a_n, c) = 1$. This is a unique expression for the elements of H, up to multiplication by -1.

Once we have an expression for hyperplanes in terms of integers, it's easy to get a hyperplane arrangement in F_q^n from a hyperplane arrangement $\mathcal{A}$ in R^n: Just interpret these integers as elements of F_q. The resulting hyperplane arrangement is denoted $\mathcal{A}(\mathsf{F}_q)$. When passing to a finite field, some information might naturally be lost; the discussion after Example 6.4 shows exactly this happening. Nevertheless, in a mathematician's sense of the the phrase, this almost never happens, at least for the lattice structure.

Proposition 6.14. *For all but finitely many primes p, $L_{\mathcal{A}} \cong L_{\mathcal{A}(\mathbb{F}_{p^r})}$.*

This proof uses a bit of linear algebra; see Appendix A.2 for a review.

Proof of Proposition 6.14. Let $\mathcal{A} = \{H_1, \dots, H_m\}$ with

$$H_i = \{x \in \mathbb{R}^n : a_{i,1}x_1 + \cdots + a_{i,n}x_n = c_i\}$$

for every $i \in [m]$. Then $H_{i_1} \cap \cdots \cap H_{i_k}$ is nonempty if and only if, setting $A = (a_{i_r,j})_{1\le r\le k, 1\le j\le n}$ and $(c_{i_r})_{1\le r\le k}$, there is a solution to the equation $Ax = c$. This is true if and only if $\operatorname{rank}(A|c) = \operatorname{rank}(A)$, in which case the set of solutions has dimension $n - \operatorname{rank}(A)$. The information about the ranks of the row-submatrices of A and $[A|c]$ can be obtained solely from the information on which minors of $[A|c]$ are nonzero. This means that the lattice structure is preserved if the set of vanishing minors does not change.

There are only finitely many primes that divide the minors of $[A|c]$; if p is any other prime, then the lattices $L_{\mathcal{A}}$ and $L_{\mathcal{A}(\mathbb{F}_{p^r})}$ are the same. □

We can see in a more rigorous way from this proof why graphical arrangements are independent of the base field: the matrix A has exactly one 1 and one -1 in every row. And:

Problem 6.5. Show that, in a matrix with exactly one 1 and one -1 in each row, every minor has value 0, 1, or -1.

But here is the upshot.

Theorem 6.15. *For any rational hyperplane arrangement $\mathcal{A} = \{H_1, \dots, H_m\}$ in $\mathbb{R}^n$, there is a natural number N such that for every $p > N$ and $q = p^r$,*

$$\chi_{\mathcal{A}}(q) = \left|\mathbb{F}_q^n \setminus \bigcup_{H \in \mathcal{A}(\mathbb{F}_q)} H\right|.$$

Proof. Take N as given in Proposition 6.14. For any such q, we have $L_{\mathcal{A}} = L_{\mathcal{A}(\mathbb{F}_q)}$. From Definition 6.12, the characteristic polynomials are the same. Now apply Proposition 6.13. □

Techniques like this that reduce a continuous combinatorial problem to one over a finite field are generally referred to as falling under the *finite field method*. Let's continue to study hyperplane arrangements, this time in a way that we cannot study those over finite fields.

For real hyperplane arrangements, the hyperplanes divide the ambient space into several different regions—something that hyperplanes over a finite field do not. Instead of counting the number of points in $\left|\mathbb{R}^n \setminus \bigcup_{H \in \mathcal{A}} H\right|$, let's now count the number of connected components.

Definition 6.16. Given a hyperplane arrangement $\mathcal{A}$ in $\mathbb{R}^n$, we denote by $r(\mathcal{A})$ the number of connected components of $\mathbb{R}^n \setminus \bigcup_{H \in \mathcal{A}} H$ (also called the *regions* of $\mathcal{A}$).

We might also count the number of bounded regions of a hyperplane arrangement. Here we have to be a little careful: It's possible that there are *no* bounded regions. (For example, if $\mathcal{A}$ is a collection of parallel hyperplanes.)

Exercise 6.6. Show that if a hyperplane arrangement $\mathcal{A} \subseteq \mathbb{R}^n$ has a bounded region, then the normal vectors to the hyperplanes in $\mathcal{A}$ span all of $\mathbb{R}^n$.

For this reason, we usually consider the number of *relatively* bounded regions.

Definition 6.17. Suppose that $\mathcal{A}$ is a hyperplane arrangement in $\mathbb{R}^n$ and W is the subspace of $\mathbb{R}^n$ spanned by the normal vectors to the hyperplanes in $\mathcal{A}$. The *rank* of $\mathcal{A}$ is the dimension of W. A *relatively bounded region* of $\mathcal{A}$ is a bounded connected component in $W \setminus \bigcup_{H \in \mathcal{A}} H$; we denote the number of relatively bounded components by $b(\mathcal{A})$.

Exercise 6.7. Taking $\mathcal{A}$ and W as in the previous definition, show that $L_{\mathcal{A}}$ and $L_{\mathcal{A}|_W}$ are the same poset. (Here, $\mathcal{A}|_W$ denotes the hyperplane arrangement in $\mathbb{R}^{\mathrm{rank}(\mathcal{A})} \cong W$ obtained by restricting the hyperplanes in $\mathcal{A}$ to W.)

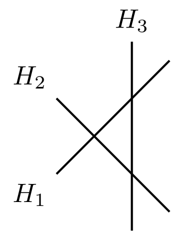
This information, too, is bound up in the characteristic polynomial.

Theorem 6.18 (Zaslavsky, 1975). *If $\mathcal{A}$ is a hyperplane arrangement in $\mathbb{R}^n$, then*

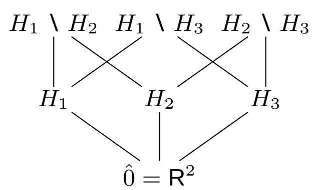
$$r(\mathcal{A}) = (-1)^n \chi_{\mathcal{A}}(-1)$$
$$b(\mathcal{A}) = (-1)^{\mathrm{rank}(\mathcal{A})} \chi_{\mathcal{A}}(1).$$

One proof of this theorem relies on an extension of the deletion-contraction method for graphs; for hyperplanes, this becomes the "deletion-restriction" method. For details, see Section 3.11.2 of *Enumerative Combinatorics 1* or Lecture 2 of Richard Stanley's notes on hyperplane arrangements.

Example 6.19. Consider the hyperplane arrangement $\mathcal{A}$ in $\mathbb{R}^2$ pictured below:



Its intersection poset looks like this:



We can calculate the Möbius values directly: Either by the recursive formula or by noticing that this poset is a subposet of the Boolean lattice B_3. Either way, $\mu_{L_{\mathcal{A}}}(\hat{0}, S) = -1$ if S is in the middle row and $\mu_{L_{\mathcal{A}}}(\hat{0}, S) = 1$ if S is in the top row. Looking at Definition 6.12, we find that

$$\chi_{\mathcal{A}}(t) = t^2 - 3t + 3.$$

According to Theorem 6.18, the arrangement $\mathcal{A}$ should have $(-1)^2 \chi_{\mathcal{A}}(-1) = 7$ different regions, which it does. You can see visually that $\mathrm{rank}(\mathcal{A}) = 2$, so relatively bounded regions are simply bounded regions. The arrangement should have $(-1)^2 \chi_{\mathcal{A}}(1) = 1$ of them, which it does.

When specializing Theorem 6.18 to graphical arrangements, we get something quite nice.

Corollary 6.20 (Stanley, 1973). *The number of acyclic orientations of a graph G with n vertices is $(-1)^n \chi_G(-1)$.*

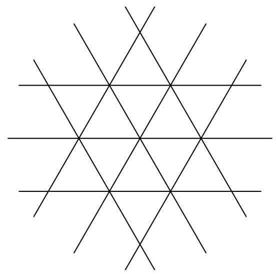
Exercise 6.8. Prove Corollary 6.20 by finding a bijection between the regions of $\mathcal{A}$ and the acyclic orientations of G.

6.4. the catalan arrangement

Definition 6.21. The *Catalan arrangement* 😺_n in $\mathbb{R}^n$ consists of the hyperplanes whose defining equations are $x_i - x_j = c$ with $c \in \{0, \pm 1\}$.

The normal vectors to the hyperplanes in 😺_n span the subspace of $\mathbb{R}^n$ that is orthogonal to the all-ones vector $\mathbf{1}$, so the rank of 😺_n is $n-1$.

Example 6.22. We can project 😺_3 onto the orthogonal subspace of $\mathbf{1}^{?}$ to get a visualization in two dimensions. It looks like this:



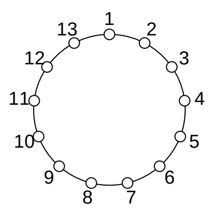
At the moment, it's mysterious why this particular arrangement is called "Catalan." We'll see why this is after we calculate its characteristic polynomial.

Proposition 6.23. *$\chi_{\text{😺}_n}(t) = t\,(t-n-1)^{\underline{n-1}}$, where $x^{\underline{k}}$ is the kth falling power of x.*

Proof. We use the finite field method. Fix a prime p such that Theorem 6.15 holds, so that

$$\chi_{\text{😺}_n}(p) = \#\big\{(x_1, \ldots, x_n) \in \mathbb{F}_p^n : x_i - x_j \neq 0, \pm 1 \text{ for every } i, j \in [n]\big\}.$$

Let's make this more combinatorial. Imagine a necklace with p different (distinguishable) beads. We want to count the number of ways to choose n beads in a particular order from this necklace so that no two are of these beads are adjacent. We have something that looks like this (in the special case $p = 13$, but you can surely imagine it with larger p):



Let's call a sequence of beads where no two are adjacent *separated*. We can form equivalence classes of separated sequences by rotation; every equivalence class has exactly p elements, exactly one of which has bead 1 as its first element.

Once we know that bead 1 is in the separated sequence, the remaining beads are simply a separated sequence of length $n-1$ in the interval $[3, p-1]$. Each separated *set* in $[3, p-1]$ corresponds to exactly $(n-1)!$ different separated sequences. Now we're almost done: Each separated set in $[3, p-1]$ is, perhaps uninterestingly, a separated set in $[2, p]$ that does not contain 2 or p. Such sets exactly correspond to compositions of $p-n$ into n parts: the beads in the set separate the $p-n$ unchosen beads into the different components of the composition. And we know from the twelvefold way that there are exactly $\binom{p-n-1}{n-1}$ compositions of $p-n$ into n parts.

Altogether, then, we have

$$\chi_{\text{😺}_n}(p) = p(n-1)!\binom{p-n-1}{n-1} = p\,(p-n-1)^{\underline{n-1}}.$$

Since this is true for infinitely many integers p, it is an equality of polynomials, which proves the theorem. $\square$

Corollary 6.24. *The number of relatively bounded regions of the Catalan arrangement* 😺$_n$ *is* $n!C_n$.

Proof. Apply Theorem 6.18 to get

$$r(😺_n) = (n+2)(n+3)\cdots(2n) = n!\frac{1}{n+1}\binom{2n}{n}. \qquad \square$$

Exercise 6.9. The *braid arrangement* in R^n contains the hyperplanes defined by the equations $x_i - x_j = 0$. (It is the graphical arrangement of the complete graph K_n.) Determine its characteristic polynomial and show that it has $n!$ regions.

It's also possible to see that the braid arrangement has $n!$ regions without recourse to Zaslavsky's theorem. The region a point $x = (x_1, \ldots, x_n)$ lies entirely depends on which side of each hyperplane that the point lies on. For the hyperplane determined by $x_i - x_j = 0$, all the points on one side satisfy $x_i < x_j$; all the points on the other satisfy $x_i > x_j$. The region that x lies in is therefore entirely determined by the relative order of its coordinates. There are $n!$ total orders of n elements, and each of these orders corresponds to a region, so there are $n!$ regions.

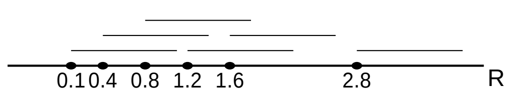
Notice that the Catalan arrangement contains within it the braid arrangement. Each region of the braid arrangement is called a *chamber* of the Catalan arrangement. The chamber in which $x_1 < x_2 < \cdots < x_n$ is called the *fundamental chamber*. The Catalan arrangement is the same in each chamber (because it is invariant under permutation of coordinates), so Corollary 6.24 can be equivalently formulated as the statement that the fundamental chamber of 😺$_n$ contains C_n regions. Our next goal is to obtain some enumerative insight into this fact.

Problem 6.10. The *Shi arrangement* in R^n contains the hyperplanes defined by the equations $x_i - x_j = c$ with $c \in \{0, 1\}$. Determine its characteristic polynomial and the number of regions into which it divides R^n.

6.5. interval orders

Suppose that $x = (x_1, \ldots, x_n)$ is a point in the fundamental chamber of 😺$_n$; that is, $x_1 < x_2 < \cdots < x_n$. To determine the region of 😺$_n$ in which x lies, we need to determine on which side of the hyperplanes it is. We know that $x_i - x_j > 0$ whenever $i > j$ for any point in the fundamental chamber, so the only relevant hyperplanes are the ones defined by $x_i - x_j = 1$. The region that x lies in, then, is determined by which pairs of i and j with $i > j$ satisfy $x_i - x_j < 1$.

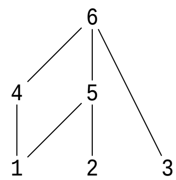
To visualize this, let's plot $x_1, \ldots, x_n$ on the real number line. We can draw unit intervals above each point to visualize the inequality; the region of x is determined by which intervals overlap. For example, for the point $(0.1, 0.4, 0.8, 1.2, 1.6, 2.8)$, the picture looks like this:



We can encode these overlaps (or lack thereof) into a poset.

Definition 6.25. Let U be a collection of intervals in R. The *interval order* obtained from U is a poset whose elements are the intervals of U with the partial order $I \prec J$ if the interval I lies entirely to the left of J (that is, the right endpoint of I precedes the left endpoint of J). A *unit interval order* (or *semiorder*) is an interval order where every interval in U has length 1.

In the collection of unit intervals above, let's label the intervals by $1, 2, 3, 4, 5, 6$ according to their left endpoint. The unit interval order derived from the collection of intervals above is



Much like Kuratowski's theorem, which characterizes just two graphs as obstructions to planarity—and many other theorems on forbidden subgraphs—interval orders have a characterization via forbidden subconfigurations. We say that a poset P has another poset Q as an induced subposet if there is a set $X \subseteq P$ such that $(X, \preccurlyeq|_X)$ is isomorphic to Q.

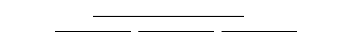
Theorem 6.26 (Scott-Suppes, 1958). *A poset is a unit interval order if and only if it does not contain either of the following as induced subposets:*

Theorem 6.27 (Fishburn, 1970). *A poset is a unit interval order if and only if it does not contain*  *as an induced subposet.*

See this 1993 paper of Bogart for a different proof than Fishburn's original.

These theorems indicate that not every unit interval order is an interval order. This is indeed the case.

Exercise 6.11. Explain why the interval order induced by this collection of intervals



cannot be realized as a unit interval order.

You can check, if you want, that the number of unlabelled Hasse diagrams of unit interval orders with 1, 2, 3, and 4 elements is 1, 2, 5, and 14, respectively—a very suspicious sequence.

Theorem 6.28. *There are C_n unlabelled Hasse diagrams of unit interval orders with n elements.*

How to prove this? One way is to introduce something new: Given a poset $(P, \preccurlyeq)$ with elements $a_1, \dots, a_n$, its *incidence matrix* $I_P = (m_{i,j})$ is the $n \times n$ matrix with $m_{i,j} = 1$ if $a_i \prec a_j$ and $m_{i,j} = 0$ otherwise.

Problem 6.12. Prove that a finite poset P is a unit interval order if and only if its elements can be ordered such that, in the incidence matrix $I_P = (m_{i,j})$, if $m_{i,j} = 1$, then $m_{r,s} = 1$ whenever $r \leq i$ and $s \geq j$. (Hint: Use Theorem 6.26.) Moreover, each unit interval order corresponds to exactly one such matrix, and each strictly upper-triangular $(0, 1)$-matrix (all 0's on the diagonal) of this form corresponds to a unit interval order.

Once we have this result, it's not so hard to count the number of unlabelled unit interval orders by counting the number of these matrices. The matrices look similar to this one:

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

If we draw the dividing line,

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

then rotate by 45°, we get a Dyck path with $2n$ steps. Problem 6.12 guarantees that this a bijection. Notice the similarity to the proof of Proposition 5.28.

Question 6.29. Can you find a "natural" bijection (whatever this might mean) between the set of noncrossing partitions of $[n]$ and the set of unlabelled unit interval orders with n elements?

This still doesn't quite explain why the number of regions in the fundamental chamber of 😺$_n$ is C_n, since those are a particular class of *labelled* unit interval orders. There's one missing step:

Exercise 6.13. Suppose that $U = \{I_1, \ldots, I_n\}$ is a collection of unit intervals in $\mathbb{R}$ such that the left endpoint of I_i is less than the left endpoint of I_j if and only if $i < j$. Show that the incidence matrix of the interval order induced by U satisfies the properties of Problem 6.12. Then use this along with the other results in this section to show that the number of regions in 😺$_n$ is C_n.

6.6. complex hyperplane arrangements

Let's take this up a notch and consider hyperplane arrangements over the field of complex numbers. Hyperplanes in $\mathbb{C}^n$ are defined just as they are over any field: the solutions to an equation $a_1x_1 + \cdots + a_nx_n = c$ with $a_i, c \in \mathbb{C}$.

These behave fairly differently from hyperplane arrangements over $\mathbb{R}$. It no longer is much use to count the number of connected components of a hyperplane arrangement, because there is always only 1. To understand why, consider a hyperplane arrangement in $\mathbb{C}^1$. The "hyperplanes" in this case are just affine subspaces of $\mathbb{C}^1$ of dimension 0: points. Removing a finite number of points from $\mathbb{C}^1$ doesn't disconnect it, so $\mathbb{C}^1 \setminus \mathcal{A}$ remains connected.

In general, the map $a + bi \mapsto (a, b)$ extends to a homeomorphism from $\mathbb{C}^n$ to $\mathbb{R}^{2n}$ that takes affine subspaces in $\mathbb{C}^n$ of dimension k to affine subspaces of $\mathbb{R}^{2n}$ of dimension $2(k-1)$. Subtracting a finite number of affine subspaces of $\mathbb{R}^{2n}$, each with dimension at most $2n-2$, keeps the remaining space connected.

So counting regions is out. Instead, we'll turn to something more topological in nature. Every topological space X has an associated *cohomology ring* $H^*(X)$; it is a graded ring, meaning that there is a canonical decomposition of the underlying abelian group into a direct sum

$$H^*(X) = H^0(X) \oplus \cdots \oplus H^n(X)$$

in a way that the multiplication respects: If $x \in H^r(X)$ and $y \in H^s(X)$, then $xy \in H^{r+s}(X)$. (If X is some kind of pretty crazy space, the direct sum decomposition may not be finite.) Each of the groups $H^i(X)$ is actually a vector space over $\mathbb{C}$ and therefore isomorphic to $\mathbb{C}^{\beta_i}$ for some $\beta_i \in \mathbb{N}$ called the ith *Betti number* of X. Betti numbers are useful algebraic invariants of topological spaces, so they're sometimes packaged into a single invariant called the *Poincaré polynomial* $\mathrm{Poin}_X(t)$. The Poincaré polynomial is one of the more important algebraic invariants of a space.

That's a lot of algebra. Here's the connection to hyperplane arrangements. Given an arrangement $\mathcal{A}$ in $\mathbb{C}^n$, let $C_{\mathcal{A}}$ denote the topological space $\mathbb{C}^n \setminus \bigcup_{H \in \mathcal{A}} H$. The characteristic polynomial, it turns out, is really the Poincaré polynomial in disguise.

Theorem 6.30 (Orlik-Solomon, 1980). *For any hyperplane arrangement $\mathcal{A}$ in $\mathbb{C}^n$,*

$$\mathrm{Poin}_{C_{\mathcal{A}}}(t) = t^n \chi_{\mathcal{A}}(t^{-1}).$$

To a certain extent, this explains why the characteristic polynomial is so important in studying hyperplane arrangements—the Poincaré polynomial is important, and the characteristic polynomial is its combinatorial incarnation. Of course, this doesn't explain why the Poincaré polynomial is important. That's best left for a course in algebraic topology.

Corollary 6.31. *If $\mathcal{A}$ is a hyperplane arrangement in* $\mathbb{R}^n$ *and* $\mathcal{A}_{\mathbb{C}}$ *is its "complexification" (that is, the collection of hyperplanes in* $\mathbb{C}^n$ *defined by the same equations as in* $\mathcal{A}$*), then the number of regions in* $\mathcal{A}$ *is equal to* $\mathrm{Poin}_{\mathcal{A}_{\mathbb{C}}}(-1)$.

Proof. Apply Theorem 6.18. □

Orlik and Solomon gave a completely combinatorial description of the cohomology ring of $C_{\mathcal{A}}$. If you're not interested in this, just skip the next section.

6.7. the orlik-solomon algebra

Orlik and Solomon, in 1980, gave a combinatorial basis for the cohomology ring of any complex hyperplane arrangement. Here's how it works. Start with a hyperplane arrangement $\mathcal{A} = \{H_1, \dots, H_N\}$ in $\mathbb{C}^n$. We start with N generators $\alpha_1, \dots, \alpha_N$. The product of two generators, denoted $\wedge$, satisfies the following rules:

1. $\alpha_i \wedge \alpha_i = 0$ for all $i \in [N]$.
2. $\alpha_i \wedge \alpha_j = -\alpha_j \wedge \alpha_i$ if $i \neq j$.
3. If $\bigcap_{r=1}^{m} H_{i_r} = \emptyset$, then $\alpha_{i_1} \wedge \cdots \wedge \alpha_{i_m} = 0$.
4. If $\bigcap_{r=1}^{m} H_{i_r} = \emptyset$ and the normal vectors to $H_{i_1}, \dots, H_{i_m}$ form a minimal linearly dependent set, then:

$$\sum_{r=1}^{m} \alpha_{i_1} \wedge \cdots \wedge \alpha_{i_{r-1}} \wedge \alpha_{i_{r+1}} \wedge \cdots \wedge \alpha_{i_m} = 0.$$

The $\mathbb{C}$-algebra generated by $\alpha_1, \dots, \alpha_N$ modulo these relations is the Orlik-Solomon algebra, denoted $OS_{\mathcal{A}}$. The central result is this:

Theorem 6.32 (Orlik-Solomon, 1980). *The algebra* $OS_{\mathcal{A}}$ *is isomorphic to the cohomology ring of* $C_{\mathcal{A}} = \mathbb{C}^n \setminus \bigcup_{H \in \mathcal{A}} H$.

If you know anything about de Rham cohomology, the isomorphism to the de Rham cohomology ring is given by

$$\alpha_i \mapsto d \log\big(h_i(x) - c_i\big) = \frac{d\big(h_i(x) - c_i\big)}{h_i(x) - c_i},$$

where H_i is defined by the equation $h_i(x) = c_i$.

6.8. new formulas for the characteristic polynomial

Let's get back to real hyperplane arrangements. By this point, you may not be surprised that the characteristic polynomial has been studied from various points of view. Here is another.

Definition 6.33. Let $\mathcal{A} = \{H_1, \dots, H_N\}$ be a hyperplane arrangement in $\mathbb{R}^n$. A subset $I \subseteq [N]$ is called *central* if $\bigcap_{i \in I} H_i \neq \emptyset$. The *rank* of I is $\mathrm{rank}(I) = \dim\big(\mathrm{span}(v_i : i \in I)\big)$. By convention, the empty set is central.

Whitney's theorem gives a formula for the characteristic polynomial in terms of the central subsets of a hyperplane arrangement.

Theorem 6.34 (Whitney's theorem). *If $\mathcal{A}$ is an arrangement of N hyperplanes in $\mathbb{R}^n$, then*

$$\chi_{\mathcal{A}}(t) = \sum_{\substack{I \subseteq [N] \\ I \text{ central}}} (-1)^{|I|} t^{n-\operatorname{rank}(I)}.$$

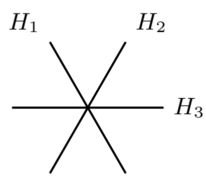
We'll prove this later for now, see how it transforms region-counting:

Corollary 6.35. *If $\mathcal{A}$ is an arrangement of N hyperplanes in $\mathbb{R}^n$, then*

$$r(\mathcal{A}) = \sum_{\substack{I \subseteq [N] \\ I \text{ central}}} (-1)^{|I|-\operatorname{rank}(I)} \qquad \text{and} \qquad b(\mathcal{A}) = \sum_{\substack{I \subseteq [N] \\ I \text{ central}}} (-1)^{|I|}.$$

Proof. Apply Theorem 6.18. □

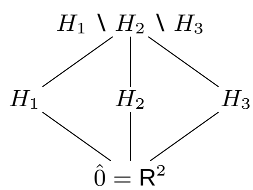
Example 6.36. Consider the following arrangement in $\mathbb{R}^2$.



Every collection of hyperplanes intersects, so every collection is central. Following Theorem 6.34, each of the singleton subsets of $[3]$ contributes $-t$ to the characteristic polynomial; the pairs each contribute 1; the whole set contributes -1; and the empty set contributes t^2. So

$$\chi_{\mathcal{A}}(t) = t^2 - 3t + 3 - 1 = t^2 - 3t + 2.$$

You can check that this is indeed the characteristic polynomial of $\mathcal{A}$ by using its intersection lattice:



We can see from Example 6.36 that some of the terms in the sum over central subsets might cancel with each other; in other words, the formula of Theorem 6.34 is not *subtraction-free*.

Whitney's theorem can be derived as a corollary of a general theorem about Möbius functions on lattices.

Theorem 6.37 (Crosscut theorem). *Suppose that L is a finite lattice and X is a subset of elements of L such that*

1. *$\hat{0} \notin X$ and*
2. *if $y \in L \setminus \{\hat{0}\}$, then there is an $x \in X$ for which $y \le x$.*

If n_k denotes the number of k-element subsets of X whose join is $\hat{1}$, then

$$\mu_L(\hat{0}, \hat{1}) = \sum_{k=0}^{|X|} (-1)^k n_k.$$

An element of a lattice is called an *atom* if it covers $\hat{0}$. For intuition, notice that any set X that satisfies the crosscut conditions must include every atom. In fact, X satisfies the crosscut conditions *if and only if* it contains every atom.

Example 6.38. In the Boolean lattice B_n, the singleton sets are the atoms. Take X to be the collection of atoms; how many ways are there to write $\hat{1} = [n]$ as the join of k singleton sets? Exactly one way, if $k = n$, and exactly 0 ways if $k < n$. So $\mu_{B_n}(\hat{0}, \hat{1}) = (-1)^n$.

Exercise 6.14. Use the crosscut theorem to calculate $\mu_{L_q(n)}(\hat{0}, \hat{1})$, where $L_q(n)$ is the poset of subspaces of $\mathbb{F}_q^n$.

For a proof, see *Enumerative Combinatorics*, Section 3.9. Stanley's proof is interesting: He defines a new algebra structure using the lattice and its Möbius function and it then simplifies to a few calculations. I encourage you to take a look.

As for us, we'll use the crosscut theorem to prove Whitney's.

Proof of Theorem 6.34 from crosscut. For each $S \in L_A$, define $L(S) = \{T \in L_A : T \preccurlyeq S\}$. This is a lattice. (Geometrically, it's the lattice of subspaces in L that contain S.) The collection of hyperplanes $\{H \in \mathcal{A} : H \supseteq S\}$ is the collection of atoms of $L(S)$; applying the crosscut theorem, we get

$$\mu_{L_\mathcal{A}}(\hat{0}, S) = \mu_{L(S)}(\hat{0}, \hat{1}) = \sum_k (-1)^k n_k = \sum_k \sum_{\substack{H \in \binom{\mathcal{A}}{k} \\ \bigcap H = S}} (-1)^k = \sum_{\substack{H \subseteq A \\ \bigcap H = S}} (-1)^{|H|}.$$

Now insert this evaluation into the definition $\chi_A(t) = \sum_{S \in L_\mathcal{A}} \mu_{L_\mathcal{A}}(\hat{0}, S) t^{\dim S}$ and use the fact that $\dim(S) = n - \operatorname{rank}(S)$. The resulting formula is Theorem 6.34. □

Whitney's theorem significantly reduces the number of terms required to calculate the characteristic polynomial of a hyperplane arrangement. But you can ask: Is it possible to do any better? We saw already that some of the terms in the some might cancel, so the formula may still be somewhat redundant. (In fact, it might be significantly redundant; the hyperplane arrangement in $\mathbb{R}^2$ that consists of 1000 lines through the origin will have a *lot* of cancellation.) Reducing the redundancy is exactly the goal of the No Broken Circuits theorem.

Definition 6.39. Let $\mathcal{A} = \{H_1, \ldots, H_N\}$ be a hyperplane arrangement in $\mathbb{R}^n$, and let v_i be a normal vector to H_i for each $i \in [N]$. A central subset $I \subseteq [N]$ is

- *independent* if $\{v_i\}_{i \in I}$ is linearly independent;
- a *circuit* if $\{v_i\}_{i \in I}$ is linearly dependent but any proper subset of it is independent;
- a *broken circuit* if it can be written as $I = C \setminus \{m\}$ where C is a circuit and $m = \min C$.

A subset $I \subseteq [N]$ is called an NBC subset (for "no broken circuit") if it is central and it does not contain any broken circuits as subsets.

Theorem 6.40 (No Broken Circuits theorem). *If $\mathcal{A} = \{H_1, \ldots, H_N\}$ is a hyperplane arrangement in $\mathbb{R}^n$, then*

$$\chi_A(t) = (-1)^n \sum_{\substack{I \subseteq [N] \\ I \text{ is NBC}}} t^{n-|I|}.$$

In other words,

$$\operatorname{Poin}_A(t) = \sum_{\substack{I \subseteq [N] \\ I \text{ is NBC}}} t^{|I|}.$$

And here we have a completely subtraction-free formula!

Proof of Theorem 6.40. From Whitney's theorem, we have

$$\chi_A(t) = \sum_{\substack{I \subseteq [N] \\ I \text{ central}}} (-1)^{|I|} t^{n-\operatorname{rank}(I)}.$$

If we can construct a sign-reversing and rank-preserving involution on central subsets that contain a broken circuit, this will show that all the terms except the ones corresponding to NBC subsets cancel.

So, here's the map: If given a set $I \subseteq [N]$, let $m(I)$ be the greatest element of $[N]$ such that I contains a broken circuit $C \setminus \{m\}$ (the element m need not be in I). We then define $\varphi(I) = I \triangle \{m\}$; that is, $\varphi(I) = I \cup \{m\}$ if $m \notin I$ and $\varphi(I) = I \setminus \{m\}$ if $m \in I$. Since m is a member of a circuit, φ preserves rank; it changes the number of elements in I by 1, so it also reverses sign.

To show that φ is an involution, we need to prove that $m(I) = m\big(\varphi(I)\big)$. Suppose that $m = m(I) \in I$ (the other case is similar). Since every broken circuit of $\varphi(I)$ is a broken circuit of I, we have $m\big(\varphi(I)\big) \leq m(I)$. On the other hand, if $C \setminus \{r\}$ is a broken circuit of I with $r \geq m$, then every element of C is greater than m; therefore $C \setminus \{r\} \subseteq \varphi(I)$. It follows that $m(I) |> m\big(\varphi(I)\big)$; so they are equal. Since $m(I) = m\big(\varphi(I)\big)$, we have $\varphi^2(I) = I$, which finishes the proof. □

Exercise 6.15. To calculate the characteristic polynomial of the braid arrangement, we need to count NBC subsets of the edges of K_n, the complete graph on the vertex set $[n]$. So order the edges of K_n lexicographically: If ij and uv are edges with $i < j$ and $u < v$, then $ij \prec uv$ if either (i) $i < u$ or (ii) $i = u$ and $j < v$. What is an NBC on the set of edges of K_n with respect to the lexicographic ordering?

7. matroids

7.1. what are they?

Let's start, as usual in a new chapter, with some definitions.

Definition 7.1. A collection B of nonempty subsets of a given set E satisfies the *exchange axiom* if, for every pair of sets $I, J \in B$ and every $i \in I$, there is a $j \in J$ so that $(I \setminus \{i\}) \cup \{j\} \in B$. A *matroid* of rank d is a pair $(E, \mathcal{B})$, where E is a (finite) set, the so-called *ground set* of M, and $\mathcal{B}$ is a collection of subsets of E, each with cardinality d, that satisfies the exchange axiom. The elements of $\mathcal{B}$ are called the *bases* of M. A subset $I \subseteq E$ is called *independent* if it is contained in some basis and *dependent* otherwise. A minimally dependent set is called a *circuit*.

At first sight, the exchange axiom seems a contrived condition. The best way to think of it, to my mind, is to consider it an abstraction of linear independence. If you have a vector space V and two bases $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$, then whenever you remove a vector u_i from the first basis, you can find a vector v_j in the second basis to replace it and maintain the linear independence and spanning properties. Seen from a mile away, this is the exchange axiom.

Exercise 7.1. Suppose that $M = (E, \mathcal{B})$ is a matroid and I, J are independent sets in M with $|J| > |I|$. Show I and J satisfy the *augmentation property*; that is, there exists an element $j \in J$ such that $I \cup \{j\}$ is independent.

Matroids can be defined in many different ways; fundamentally, it is a collection of subsets of a given set that generalizes independence. The augmentation property provides a different way to characterize matroids.

Exercise 7.2. Suppose that $\mathcal{I}$ is a collection of subsets of a finite ground set E such that

- $\emptyset \in \mathcal{I}$,
- if $I \in \mathcal{I}$ and $J \subseteq I$, then $J \in \mathcal{I}$; and
- if $I, J \in \mathcal{I}$ and $|J| > |I|$, then I and J satisfy the augmentation property.

We call an element $I \in \mathcal{I}$ a *basis* if it is maximal: $I \cup \{x\} \notin \mathcal{I}$ for every $x \in E \setminus I$. Show that every basis of $\mathcal{I}$ has the same cardinality. Then show that the collection of bases satisfies the exchange axiom.

What have matroids got to do with what we've been doing?

Example 7.2. Let $\mathcal{A} = \{H_1, \dots, H_N\}$ be a hyperplane arrangement in F^n and $v_1, \dots, v_N$ be normal vectors to these hyperplanes, respectively. We can form a matroid M_A on the set $[N]$ by declaring $\{i_1, \dots, i_n\} \in \mathcal{B}$ if and only if $\{v_{i_1}, \dots, v_{i_n}\}$ is a basis of F^n.

A matroid M is isomorphic to another matroid N if there is a bijection $f\colon M \to N$ such that $f(I)$ is independent in N if and only if I is independent in N. A matroid M is *realizable* over a field F if there is a hyperplane arrangement over F whose matroid is isomorphic to it.

We can also think of this more algebraically. Given the set $\{v_1, \dots, v_N\}$ of normal vectors to the hyperplanes in $\mathcal{A}$, we can form the $n \times N$ matrix $[v_1 v_2 \cdots v_N]$. The set $I \subseteq [n]$ is a basis of M_A if and only if $\det(v_i)_{i \in I} \neq 0$. In fact, this has no need for a hyperplane arrangement; it depends only on a set of vectors. For this reason, realizable matroids are often defined in terms of a collection of vectors.

Exercise 7.3. Let S be a collection of points in $\mathbb{R}^n$. A subset $\{x_1, \dots, x_m\} \subseteq S$ is called *affinely dependent* if there are real numbers $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ such that $\sum_{i=1}^m \alpha_i = 0$ and $\sum_{i=1}^m \alpha_i x_i = 0$. The matroid M_S has S as a ground set and all affinely *in*dependent sets as, well, independent sets.

(i) The affine hull of S is the smallest affine subspace (translation of a linear subspace) that contains S. Show that S is affinely independent if and only if the affine hull of $S \setminus \{x\}$ is strictly contained in the affine hull of S for every $x \in S$. (This part is not necessary for the following parts, but it provides a good visualization of affine (in)dependence. For another description, see this post.)

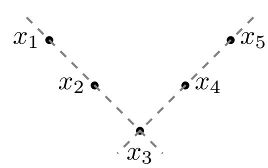
(ii) Check that this is a matroid.

(iii) Show that the collection $\{x_1, \dots, x_m\}$, with $x_i = (x_{i,1}, \dots, x_{i,n})$, is affinely independent if and only if the collection of vectors $\left\{(x_{i,1}, \dots, x_{i,n}, 1)\right\}_{i=1}^m \subset \mathbb{R}^{n+1}$ is linearly independent.

(iv) A matroid M is called *affinely realizable* if there is a point set $S \subset \mathbb{R}^n$ such that $M \cong M_S$. Show that a matroid is affinely realizable if and only if it is realizable over $\mathbb{R}$.

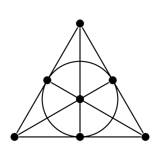
One nice thing about affine realizability is that it's easier to draw points than it is to draw vectors.

Example 7.3. Consider the following collection S of points in $\mathbb{R}^2$:



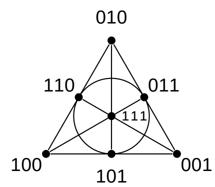
The dashed lines indicate the two affinely dependent triples $\{x_1, x_2, x_3\}$ and $\{x_3, x_4, x_5\}$. The bases of M_S are all triples besides these two. The circuits of M_S are $\{1,2,3\}$ and $\{3,4,5\}$, but also $\{1,2,4,5\}$, which is dependent but contains no dependent subset.

Example 7.4. Not every matroid is realizable. Consider the following diagram:



This diagram, which is just a visualization and *not* an affine realization diagram, represents the matroid M whose ground set is the seven vertices in this diagram and whose bases are all triples besides the ones contained in a single line (the circle is a line for the purposes of this visualization). You can check that this is, in fact, a matroid. However, it cannot be realized over $\mathbb{R}$! Since every basis has 3 elements, if it could be realized over $\mathbb{R}$, it must be realized as a point set in $\mathbb{R}^2$. But that's not possible—convince yourself of this.

However, it *can* be realized over $\mathbb{F}^2$, the collection of all nonzero vectors in $\mathbb{F}_2^3$, as you can see below:



Another way to realize a matroid is via graphs.

Definition 7.5. If G is a graph, the matroid M_G induced by G has ground set $E(G)$, and a subset of edges is independent in M_G if and only if it contains no cycle. (In other words, a subset of edges is independent if and only if it is a forest.) A matroid that is isomorphic to M_G for some graph G is called *graphical*.

Exercise 7.4. Suppose that G is a connected graph. Show that a set $I \subseteq E(G)$ is a basis in M_G if and only if it is a spanning tree of G; then show that I is a circuit if and only if it is a cycle (with no repeating vertices) in G.

Graphical matroids are a very special case of realizable matroids: They're realizable over *any* field.

Exercise 7.5. Let G be a graph with n vertices and F be a field.

(i) Suppose that G is the disjoint union of G_1 and G_2. Show that, if M_{G_1} and M_{G_2} are realizable over F, then G is realizable over F.

(ii) We can now assume that G is connected. For each edge $e \in E(G)$, let $v_e \in F^n$ denote the vector whose ith and jth coordinates are 1 and -1, respectively, with every other coordinate 0. (The vector v_e is a normal vector to the hyperplane in the graphical arrangement $\mathcal{A}_G$ corresponding to the edge e.) Show that, if $I \subseteq E(G)$ contains a cycle, then $\{v_e\}_{e\in I}$ is linearly dependent.

(iii) Show that, if $I \subseteq E(G)$ is a tree, then $\{v_e\}_{e\in I}$ is linearly independent. (**Hint:** If I were linearly dependent, say $\sum_{e\in I} \alpha_e v_e = 0$, then what is the coefficient α_e when e is incident to a leaf?)

(iv) Show that M_G is realizable over F.

Each matroid is equipped with a function that parallels the rank of a linear (or affine) subspace.

Definition 7.6. If $M = (E, \mathcal{B})$ is a matroid, the *rank* of a subset $I \subseteq E$, denoted $r(I)$, is the size of the largest independent subset of M contained in I.

Exercise 7.6. Show that $r(I) = \max_{B\in\mathcal{B}} |I \cap B|$.

The rank function satisfies several properties:

Proposition 7.7. *If $M = (E, \mathcal{B})$ is a matroid and r is the rank function on M, then*

1. $r(\emptyset) = 0$;
2. *for every $I \subseteq E$ and $i \in E$, we have $r(I \cup \{i\}) - r(I) \in \{0, 1\}$; and*

3. *the rank function is* submodular*: For every* $I, J \subseteq E$

$$r(I) + r(J) \geq r(I \cup J) + r(I \cap J).$$

Exercise 7.7. Prove Proposition 7.7.

It may not be clear when the submodularity inequality would be strict. Here's one example. Pick two basis elements I and J of the graphical matroid M_{K_3}. (Both I and J a pair of edges.) Then $I \cup J = K_3$ and $I \cap J$ is a single edge, so $r(I)+r(J) = 2+2 = 4$, while $r(I \cup J)+r(I \cap J) = 2+1 = 3$.

Im any case, it turns out that the rank function is enough to specify the matroid.

Exercise 7.8. Suppose that E is a finite set and $r\colon \mathcal{P} \to \mathrm{N}_0$ is a function that satisfies the three conditions of Proposition 7.7 and define $\mathcal{I} = \{I \subseteq E : r(I) = I\}$. Show that $\mathcal{I}$ forms a collection of independent sets for a matroid on E. (Hint: Use submodularity to prove: If $i, j \in E$ satisfy $r(I \cup \{i\}) = r(I)$ and $r(I \cup \{j\}) = r(I)$, then $r(I \cup \{i, j\}) = r(I)$.)

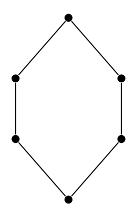
7.2. geometric lattices

Let's connect matroids to something more familiar: posets. To do that, we'll have to do a bit of an excursion; bear with me.

Definition 7.8. A lattice L is

- *atomic* if every element in L is the join of a finite number of atoms. (The element $\hat{0}$ is, by convention, the empty join.)
- *graded* if it has a rank function $\rho\colon L \to \mathrm{N}_0$ such that $\rho(\hat{0}) = 0$ and $\rho(x) = \rho(y) + 1$ whenever x m y.
- *semimodular* if it is graded with a rank function ρ that satisfies $\rho(x)+\rho(y) \geq \rho(x \wedge y)+\rho(x \vee y)$ for every $x, y \in L$.
- *geometric* if it is finite, atomic, and semimodular.

Example 7.9. The poset



is a graded lattice, but it is neither atomic nor semimodular. If we name the two atoms x and y, then the two elements of L that are not atoms cannot be written as the join of x and y. Moreover, $\rho(x) + \rho(y) = 2$, but $x \vee y = \hat{1}$ and $x \wedge y = \hat{0}$, so $\rho(x \wedge y) + \rho(x \vee y) = 0 + 3 > 2$.

Exercise 7.9. Find an atomic lattice that is not graded. (Hint: You need only three atoms.)

Exercise 7.10. Show that the Boolean lattice B_n is geometric.

Exercise 7.11. Suppose that L is an atomic lattice with exactly n atoms. Show that L is a sublattice of the Boolean lattice B_n.

In particular, Exercise 7.11 implies that any chain of length 3 or greater is not atomic (and therefore not geometric). You can derive stronger consequences: A finite atomic lattice with n generators cannot contain a path of length $n+1$ or greater; a finite atomic lattice with n generators cannot have width greater than $\binom{n}{bn/2c}$. (The *width* of a poset is the size of the largest collection of mutually incomparable elements.)

Lemma 7.10. *The intersection lattice of any (finite) central hyperplane arrangement is geometric.*

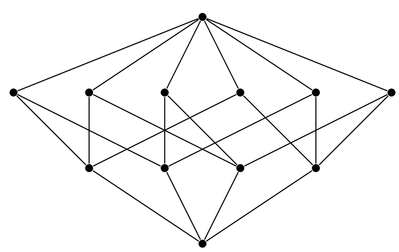
Proof. The lattice is graded with the rank function of codimension. The join of L_A is intersection and the atoms of the lattice are the hyperplanes (remember that the lattice is ordered by *reverse* inclusion). The lattice is atomic since the elements are defined by intersections of these hyperplanes. You can check that the lattice is semimodular; in fact, equality holds in the semimodularity condition. □

Here's how matroids come into the picture.

Definition 7.11. Let M be a matroid on the ground set E. A set $S \subseteq E$ is called a *k-flat* if it is a maximal subset of E with rank k; that is, if $r(S) = k$ and $r(T) > k$ for every $T \supsetneq S$. The *lattice of flats* of M, denoted $L(M)$, is the set of all flats of M ordered by inclusion.

In the independence picture, a flat corresponds to an (a□ne) subspace.

If you're like me, then you might have a hard time constructing a simple example of a geometric lattice where the semimodularity inequality is strict. Similarly, it wasn't clear to me whether the submodularity inequality of the rank function of a matroid could be strict when applied to flats. Flats give a concise way to answer both questions at the same time. The *uniform matroid* U_k^n on $[n]$ of rank k is the matroid whose basis set consists of all k-element subsets of $[n]$. The lattice $L(U_3^4)$ looks like this:



You'll show that it actually is a lattice in the next exercise; for now, simply note that you can pick two elements in the second row from the top, call them x and y, whose meet is $\hat{0}$. For these two elements, $\rho(x)+\rho(y) = 2+2 = 4$ while $\rho(x \wedge y)+\rho(x \vee y) = 0+3 = 3$. Similarly, in the matroid U_3^4, the sets $I = \{1, 2\}$ and $J = \{3, 4\}$ are flats and $r(I)+r(J) = 4$ while $r(I \cap J)+r(I \cup J) = 0+3 = 3$.

Exercise 7.12. Let's show that $L(M)$ actually is a lattice. Given a set $I \subseteq E$, let $\mathrm{cl}(I) = \{i \in E : r(I \cup \{i\}) = r(I)\}$ denote the *closure* of the set I.

1. Show that $I \subseteq \mathrm{cl}(I)$ and $\mathrm{cl}(I) = \mathrm{cl}(\mathrm{cl}\, I)$.
2. Prove that $r(\mathrm{cl}\, I) = r(I)$. (**Hint:** See Exercise 7.8.)
3. Prove that if J) $\mathrm{cl}(I)$, then $r(J) > r(I)$.
4. Show that I is a flat if and only if $\mathrm{cl}(I) = I$.
5. Prove that if I and J are flats, then $I \cap J$ is a flat.
6. What are the meet and join operations of $L(M)$?

For kicks, I'll note that the closure operator provides yet another way to define matroids.

Exercise 7.13. Prove that the closure operator satisfies the following property: If $j \in \mathrm{cl}(I \cup \{i\}) \setminus \mathrm{cl}(I)$, then $b \in \mathrm{cl}(I \cup \{j\}) \setminus \mathrm{cl}(J)$.

Exercise 7.14. Suppose that E is a finite set and $\mathrm{cl}\colon \mathcal{P}(E) \to \mathcal{P}(E)$ satisfies the following properties for every $I, J \subseteq E$:

- $I \subseteq \mathrm{cl}(I)$;
- $\mathrm{cl}(I) = \mathrm{cl}(\mathrm{cl}\, I)$;
- if $I \subseteq J$, then $\mathrm{cl}(I) \subseteq \mathrm{cl}(J)$; and

- ∘ if $j \in \mathrm{cl}(I \cup \{i\}) \setminus \mathrm{cl}(I)$, then $i \in \mathrm{cl}(I \cup \{j\}) \setminus \mathrm{cl}(I)$.

Let $\mathcal{I}$ be the collection of minimal closure-generating subsets; that is, the collection of sets $I \subseteq E$ such that if $J \subsetneq I$, then $\mathrm{cl}(J) \neq \mathrm{cl}(I)$. Show that $\mathcal{I}$ forms a collection of independent sets for a matroid on E.

Anyway:

Proposition 7.12. *Every lattice of flats is a geometric lattice.*

Exercise 7.15. Prove Proposition 7.12.

A converse is true, too:

Proposition 7.13. *Every finite geometric lattice is isomorphic to the lattice of flats of a matroid.*

Proof. Let L be a geometric lattice and E be the set of atoms of L. Let $\mathcal{I}$ be the collection of subsets of E which are join-minimal: That is, $I \in \mathcal{I}$ if and only if $\bigvee(I \setminus \{i\}) \subsetneq \bigvee I$ for every $i \in I$. We will first prove that $\mathcal{I}$ forms the collection of independent sets for a matroid M on E; then we will prove that $L(M) \cong L$.

Suppose that $I \subseteq E$ and $i \not\preccurlyeq \bigvee I$. Then $\bigvee(I \cup \{i\}) \succ \bigvee I$ in L, so $\rho\big(\bigvee(I \cup \{i\})\big) \geq \rho(\bigvee I) + 1$. On the other hand, $(\bigvee I) \wedge i = 0$; because L is geometric, we have

$$\rho\big(\bigvee(I \cup \{i\})\big) = \rho\big(\bigvee(I \cup \{i\})\big) + \rho\Big(\big(\bigvee I\big) \wedge i\Big) \leq \rho\big(\bigvee I\big) + \rho(i) = \rho\big(\bigvee I\big) + 1.$$

By induction, this implies that $r(\bigvee I) = |I|$ whenever $I \in \mathcal{I}$. Conversely, if I is not join-minimal, then it has a join-minimal subset J with $\bigvee J = \bigvee I$; then $r(\bigvee I) = r(\bigvee J) = |J| < |I|$. So $I \in \mathcal{I}$ if and only if $r(\bigvee I) = |I|$.

Clearly $\emptyset \in \mathcal{I}$. If J is not join-minimal, then any superset of J is not join-minimal. Contrapositively, if $I \in \mathcal{I}$ is join-minimal and $J \subseteq I$, then $J \in \mathcal{I}$. As for the augmentation axiom, suppose that $I, J \in \mathcal{I}$ with $|J| > |I|$. This means that $\rho(\bigvee J) > \rho(\bigvee I)$, so it cannot be the case that $j \preccurlyeq \bigvee I$ for every $j \in J$. Therefore there is some $j \in J$ such that $j \not\preccurlyeq \bigvee I$; then $\rho\big(\bigvee(I \cup \{j\})\big) = |I| + 1$, so $I \cup \{j\}$ is independent. Therefore $M := (E, \mathcal{I})$ is a matroid.

Now we show that $L(M) \cong L$. By our above work, the flats of M are precisely those sets of the form $F(I) = \{i \in E : i \preccurlyeq \bigvee I\}$. If $j \notin F(I)$, then $\bigvee(I \cup \{j\}) \succ \bigvee(I)$, so the ranks do not match; this means that $F(I)$ is a flat for every $I \subseteq E$. Conversely, any flat that contains I must also contain $F(I)$. These flats are in order-preserving bijection with the elements of L via the map $F(I) \leftrightarrow \bigvee F(I)$ precisely because L is atomic. This shows that $L(M) \cong L$. □

So we have a map φ from finite matroids to finite geometric lattices (Proposition 7.12) and a map ψ from finite geometric lattices to finite matroids. But they aren't inverses! Suppose M is a matroid with an element x such that $\{x\}$ is not an independent set; then $\psi(M) = \psi(M \setminus \{x\})$. Or, if $\{x, y\}$ is a circuit, then $\psi(M) = \psi(M \setminus \{x\})$. (Spend a moment figuring out why.) A matroid is called *simple* if it contains no dependent singletons or pairs. If we restrict to simple matroids, then φ and ψ actually are inverses.

A dependent singleton set is called a *loop*. The terminology derives from graphs: If e is a loop in a graph G, then e is not contained in any independent set in M_G. A *coloop* is a singleton that appears in every basis. For a connected graph, an edge is a coloop if and only if removing it disconnects the graph.

For hyperplane arrangements, the lattice of flats is something that we've already seen before, just in a different guise.

Proposition 7.14. *If $\mathcal{A}$ is a central hyperplane arrangement, then $L(M_A) \cong L_A$.*

Proof sketch. If $\mathcal{A} = \{H_1, \ldots, H_N\}$ and v_i is the normal vector to H_i, then the flats of M_A correspond bijectively to the intersections of hyperplanes in $\mathcal{A}$. □

7.3. matroid polytopes

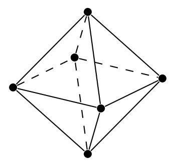
There is another way to visualize matroids that is geometric in nature rather than just in name.

Definition 7.15. Suppose M is a matroid on the ground set $[n]$ with basis set $\mathcal{B}$. We let $e_1, \ldots, e_n$ denote the standard basis of $\mathbb{R}^n$ and, for each $I \subseteq [n]$, we set $e_I = \sum_{i \in I} e_i$. The *matroid polytope* of M is

$$P_M := \mathrm{conv}(e_I : I \in \mathcal{B}).$$

If the rank of M is d, then $x_1 + \cdots + x_n = d$ for every $\mathbf{x} \in P_M$, so the dimension of P_M is at most $n-1$, though it may be less.

The polytope of the uniform matroid U_k^n is the intersection of the unit cube $[0,1]^n$ with the hyperplane $\{\mathbf{x} \in \mathbb{R}^n : x_1 + \cdots + x_n = k\}$. For U_2^4, this is the octahedron:



You can check that every every edge of this octahedron is parallel to a vector of the form $e_i - e_j$ (with $i \neq j$). In fact, this is true for any uniform matroid; you could use the definition of supporting face, for example, to check this.

It is a significant theorem that this property essentially characterizes the polytopes that arise from matroids.

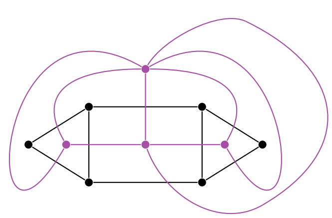
Theorem 7.16 (Gelfand, Goresky, MacPherson, Serganova 1987). *A convex polytope $P \subseteq \mathbb{R}^n$ is a matroid polytope if and only if*

1. *every vertex of P is a member of the set $\{0,1\}^n$ and*
2. *every edge of P is parallel to a vector of the form $e_i - e_j$ with $i \neq j$.*

7.4. graph duality

Let's simplify things for a moment and go back to graphs. There is a notion of duality for graphs that are embedded in the plane. Such graphs are called *plane graphs*. These are different from *planar graphs*, which simply *can* be embedded in the plane. While a planar graph is just the combinatorial data of vertices and edges, a plane graph is a planar graph together with a particular embedding of it into the plane.

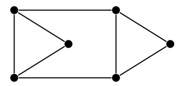
That probably seems somewhat pedantic, but we'll see the reason for the distinction in a moment. Let's see, first, how this duality works. It's actually rather simple. To take the dual of a plane graph, simply draw a new vertex at some point in each face of the graph (including the unbounded "outside face") and then draw one edge to between two new vertices through each edge that connects them. For example:



If you stare at this picture, you'll realize that each graph is the dual of the other; in other words, if you take the dual twice, you get back the same graph!

Unfortunately, the dual of a plane graph can depend on the embedding.

Exercise 7.16. Show that the dual of the plane graph



is not isomorphic to the dual of the plane graph above, even though the two base graphs are isomorphic.

Nevertheless, the dual operation on plane graphs is important; it can be used, for example, to prove that, any plane graph with v vertices, e edges, and f faces (including the unbounded face) satisfies the equality $v - e + f = 2$. (This is called Euler's formula; see the plane duality proof on David Eppstein's page of 20 different proofs (!) of Euler's formula.) The most useful basic property of the dual graph is that it interchanges faces and vertices, and the edges of the dual graph are in one-to-one correspondence with the edges of the base graph.

Let's take this duality out for a ride.

Proper edge colorings

The chromatic polynomial counts the number of proper vertex colorings, but it also counts a certain type of edge colorings.

Definition 7.17. Suppose that $G = (V, E)$ is an undirected graph and let $\vec{E}$ denote the collection of edges obtained by replacing each edge in G with two directed edges, one in each direction. We write ij for the edge directed from vertex i to vertex j. An *edge coloring* of G with colors in an abelian group Γ is a function $f \colon \vec{E} \to \Gamma$ such that

1. $f(ij) = -f(ji)$ and
2. if $i_1, \dots, i_n, i_{n+1} = i_1$ is a cycle, then $f(i_1 i_2) + f(i_2 i_3) + \cdots + f(i_n i_1) = 0$.

If further $f(ij) \neq 0$ for every $ij \in \vec{E}$, then f is called a *proper edge coloring* with values in Γ, or a proper Γ–edge coloring for short.

Definition 7.17 is quite general; the usual case is $\Gamma = \mathbb{Z}_n$, which is called an n-edge coloring, or $\Gamma = \mathbb{Z}$ or $\Gamma = \mathbb{R}$. One way to think of it is that the label $f(ij)$ is the displacement of the edge, with condition 2 guaranteeing that if you travel from one vertex and eventually land back there, then the total displacement should be 0.

Exercise 7.17. The second condition Definition 7.17 implies that the edge coloring is "independent of path." Suppose that f is an edge coloring of G with colors in Γ, and let $i_1, i_2, \dots, i_n$ and $i_1 = j_1, j_2, \dots, j_{m-1}, j_m = i_n$ be two different paths between the same vertices. Prove that $\sum_{r=1}^{n-1} f(i_r i_{r+1}) = \sum_{r=1}^{m-1} f(j_r j_{r+1})$.

Chromatic polynomials satisfy the relation $\chi_{G \sqcup H}(t) = \chi_G(t)\chi_H(t)$, which can be seen by counting the number of n-colorings of $G \cup H$ in terms of the number of n-colorings of G and H. Since $\chi_G(t)$ is divisible by t for any graph G, this means that $\chi_G(t)$ is divisible by t^c, where c is the number of components of G.

Definition 7.18. The number of components of G is denoted $\kappa(G)$. The *reduced chromatic polynomial* of G is $\widetilde{\chi}_G(t) = t^{-\kappa(G)}\chi_G(t)$.

Proposition 7.19. *The number of proper edge colorings of a graph G with colors in a finite group Γ is $\widetilde{\chi}_G(\Gamma)$*

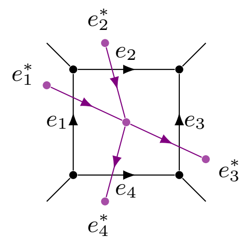
Proof. Since $\widetilde{\chi}_{G \sqcup H} = \widetilde{\chi}_G \widetilde{\chi}_H$, it suffices to prove the statement when G is connected. Every proper vertex coloring φ of G with colors in Γ induces a proper Γ–edge coloring of G defined by $(ij) = \varphi(i) - \varphi(j)$. The map $\varphi \mapsto f$ that sends vertex colorings to edge colorings is not injective;

given any $g \in \Gamma$, the vertex coloring $\varphi + g$ induces the same edge coloring as φ. On the other hand, if f is a proper edge coloring, then choosing a color for any single vertex uniquely determines the colors of every other vertex in G. So the map $\varphi \mapsto f$ is $|\Gamma|$-to-1, which shows that the number of proper edge colorings of G with colors in Γ is $\frac{1}{|\Gamma|}\chi_G(|\Gamma|) = \widetilde{\chi}_G(|\Gamma|)$. □

As before, the most common choice for Γ is $\mathbb{Z}_n$.

Flows on graphs

Now let's dualize an edge coloring. Suppose we have a plane graph G that has a proper edge coloring f with colors in Γ. We can take the plane dual of G to get another plane graph G^*; this induces an edge coloring on G^* that is not necessarily proper.



For a directed graph, the orientation of the dual edges is determined by rotating the original edge clockwise.

Let's define f^* to be the edge coloring on G^* induced by f; that is, set $f^*(e^*) = f(e)$, where e^* is the edge in G^* that crosses the edge e in G. Since f is a proper coloring, we have $f(e_1) + f(e_2) - f(e_3) - f(e_4) = 0$ (since edges e_3 and e_4 must be reversed to make a cycle). In terms of f^*, this means that $f^*(e_1^*) + f^*(e_2^*) = f^*(e_3^*) + f^*(e_4^*)$.

From the symbols, it doesn't look like much has changed, but consider the picture. For proper edges colorings, the sum over cycles is 0. When this is dualized, this turns into a sum over the edges incident to a particular vertex. In other words, the dual of a proper edge coloring is a flow.

Definition 7.20. Let G be an oriented graph with directed edge set E and Γ an abelian group. A function $f\colon E \to \Gamma$ is called a *flow* on G with values in Γ, or a Γ-flow for short, if

$$\sum_{e^+=v} f(e) = \sum_{e^-=v} f(e)$$

for every vertex v in G. The flow is called *nowhere-zero* if $f(e) \neq 0$ for every $e \in E$.

For this definition, G can be arbitrary; it need not be planar. It's easiest to think of a flow when $\Gamma = \mathbb{R}$, in which case $f(e)$ can be thought of as the amount of water flowing along edge e. The condition $\sum_{e^+=v} f(e) = \sum_{e^-=v} f(e)$ means that the amount of water flowing into each vertex is the same as the amount of water flowing out.

As before, we can think of G as being undirected with the condition that $f(ij) = -f(ji)$. The flow condition can then be restated as

$$\sum_{\substack{u \in V(G) \\ uv \in \vec{E}}} f(uv) = 0$$

for every vertex $v \in V(G)$.

We know $\widetilde{\chi}_G(n)$ is the number of proper edge colorings of G with n colors; there is a corresponding concept for nowhere-zero flows on G.

Proposition 7.21. *There is a polynomial $C_G(t)$ such that, for each finite group Γ, the number of nowhere-zero Γ-flows on G is $C_G(|\Gamma|)$.*

The polynomial in Proposition 7.21 is called the *flow polynomial* of G.

Exercise 7.18. Prove Proposition 7.21 using the deletion-contraction method.

The duality between colorings and flows can be formalized in terms of these polynomials:

Theorem 7.22 (Tutte). *If G is a connected plane graph and G^* is its dual plane graph, then $\widetilde{\chi}_G = C_{G^*}$ and $C_G = \widetilde{\chi}_{G^*}$.*

7.5. matroid duality

Duality can be extended quite broadly to the setting of matroids. That might seem odd, since non-planar graphs don't have a plane dual. But they do—they're just not graphs. Let's take a look.

Definition 7.23. Suppose that M is a matroid on E with basis set $\mathcal{B}$. The *dual* of M, denoted M^*, is the matroid on E in which $I \subseteq E$ is a basis if and only if $E \setminus I \in \mathcal{B}$.

You should be asking yourself: Is M^* a matroid? And now that you've asked, you might as well answer.

Exercise 7.19. Show that the bases of M^* satisfy the exchange axiom.

While the dual of a planar graph depends on the specific embedding of that graph, matroid duality is defined intrinsically—it depends on nothing but the matroid itself.

When we specialize to graphical matroids, we recover planar duality for graphs.

Proposition 7.24. *If G is a connected plane graph and G^* is its dual, then $M_{G^*} \cong M_G^*$.*

Proof sketch. Suppose that T is a collection of edges in G that form a spanning tree. If T^* denotes the collection of edges in G^* that do not cross any edge in T, then T^* is a spanning tree of G^*. □

The duality exhibited in Theorem 7.22 between chromatic and flow polynomials of plane graphs is extended to matroids by definition.

Definition 7.25. The *characteristic polynomial* of a matroid M is the characteristic polynomial of its lattice of flats:

$$\chi_M(t) = \sum_{x \in L(M)} \mu_{L(M)}(\hat{0}, x) t^{d - r(x)},$$

where r is the rank function of M. The *flow polynomial* of M is, by definition, $C_M(t) = \chi_{M^*}(t)$.

Since $L(M_A) = L_A$, the characteristic polynomial for matroids extends the notion for hyperplane arrangements.

We've seen a glimpse of the fact that graphical matroids have fairly nice properties; for example, they are realizable over every field. The collection of matroids that are dual to a graphical matroid, called the *cographical matroids* or *dual matroids*, shares many of its properties. The flow polynomial is one visible aspect of this; let's see another.

7.6. graphical and cographical hyperplane arrangements

Recall that, given a graph $G = (V, E)$ with n vertices, the graphical arrangement associated to it is $\mathcal{A}_G = \{H_e\}_{e \in E(G)}$, with $H_{ij} = \{x \in \mathbb{R}^n : x_i - x_j = 0\}$. Although we define $\mathcal{A}_G$ in n-dimensional space, it doesn't really "belong" in a space so big: The normal vectors to the hyperplanes in $\mathcal{A}$ span a subspace P of $\mathbb{R}^n$ with dimension $n - \kappa(G)$. (Explicitly, if $C_1, \ldots, C_m$ are the components of G and $\mathbf{1}_{C_i}$ denotes the vector that is 1 in the coordinates $j \in C_i$ and 0 elsewhere, then the subspace spanned by the normal vectors $\operatorname{span}(\mathbf{1}_{C_1}, \ldots, \mathbf{1}_{C_m})^{\perp}$.) If we consider A in this lower-dimensional space, we get that $\chi_{\mathcal{A}_G}(t) = \widetilde{\chi}_G(t)$. (This is just by dividing both sides of Theorem 6.11 by $t^{\kappa(G)}$.) Each point of P corresponds uniquely to an $\mathbb{R}$–edge coloring on G by subtracting the endpoints of each edge, and each point in $P \setminus \mathcal{A}$ is a *proper* edge coloring.

When we dualize, we should get a space of flows—the subspace of $\mathbb{R}^{|E|}$ that consists of points $f\colon E \to \mathbb{R}$ that correspond to a flow on G. (That is, points that satisfy the flow equality.) We know that the space of edge colorings of G is $\mathbb{R}^{|V| - \kappa(G)}$, and we can use this to guess the dimension of the flow space. We know from before that each edge coloring on G^* corresponds to a flow on G and vice versa. We also know that $|V(G)| = |F(G^*)|$, $|E(G)| = |E(G^*)|$, and $|F(G)| = |V(G^*)|$, as well as Euler's formula $v - e + f = 2$. From this, we find that, at least for a connected planar graph G, the flow space should have dimension

$$|V(G^*)| - 1 = |F(G)| - 1 = |E(G)| - |V(G)| + 1.$$

It turns out that this is true for any connected graph. And once you have that, it's not hard to show that the flow space of an arbitrary graph has dimension $|E(G)| - |V(G)| + \kappa(G)$. (Just use the fact that a flow on a disconnected graph is the same as a choice of flow on each component—this means that the flow space of a disjoint union of graphs is the same as the product of their flow spaces.)

Now we can define a hyperplane arrangement in the flow space of $G = (V, E)$: Take $\mathcal{A}_G^* = \{H_e\}_{e \in E}$ with

$$H_e = \{f\colon E \to \mathbb{R} : f \text{ is a flow and } f(e) = 0\}.$$

Each point in complement of $\mathcal{A}_G^*$ corresponds to a nowhere-zero flow on G. Working through the duality, one finds that $M_{\mathcal{A}_G^*} = M_{G^*}$ and

$$\chi_{\mathcal{A}_G^*}(t) = C_G(t).$$

Definition 7.26. An orientation of an undirected graph is called *totally cyclic* if every edge in the orientation is contained in a directed cycle.

Exercise 7.20. An orientation of an undirected graph is called *strongly connected* if there is a directed path from any given vertex to any other. Show that an orientation is strongly connected if and only if it is totally cyclic.

Exercise 7.21. Recall from Exercise 6.8 that the number of regions of $\mathcal{A}_G$ is the number of acyclic orientations of G. Show that the number of regions of $\mathcal{A}_G^*$ is the number of totally cyclic orientations of G. (**Hint**: A nonwhere-zero flow $f\colon \vec{E} \to \mathbb{R}$ induces an orientation on G by directing the edge ij from i to j if $f(ij) > 0$ and from j to i if $f(ji) > 0$; this orientation is the same if you choose two different flows in the same region. Now you want to show that (1) the induced orientation is totally cyclic and (2) there is an inverse function taking totally cyclic flows to the regions of $\mathcal{A}_G^*$.)

Using this result, we can apply Zaslavsky's theorem to get

Theorem 7.27 (coStanley's theorem). *The number of totally cyclic orientations of a connected planar graph G is $|\chi_{G^*}(-1)|$.*

If you substitute $\chi_{\mathcal{A}_G^*}$ or $\chi_{M_G^*}$ for χ_{G^*} (they're the same anyway), then the result holds for every graph G, not just the planar ones.

7.7. diversion: counting totally cyclic orientations

Each acyclic orientation of K_n induces a total order on the vertices, so there are exactly $n!$ acyclic orientations of K_n. The number of totally cyclic orientations, which I'll denote T_n, is more difficult. We can start by listing some numbers:

n	1	2	3	4	5
T_n	1	0	2	24	544

This sequence is a little bizarre. No matter: The next step, of course, is the run it through OEIS and see what you get. The first and only result, sequence A054946, is exactly the one we want: problem solved.

Of course, some of you in the peanut gallery might moan and groan for an explicit formula, and I'm nothing if not a people-pleaser, so let's see what we can do for you.

Proposition 7.28. *The number of totally cyclic orientations of K_n is*

$$T_n = \sum_{k=1}^{n} \sum_{\pi=(B_1,\ldots,B_k)} (-1)^{k-1} 2^{\binom{|B_1|}{2}+\cdots+\binom{|B_k|}{2}},$$

where the second sum ranges over all ordered set partitions of $[n]$ with k blocks.

And that's really the best we can do—sorry, peanut gallery. There are some asymptotics: see this paper for a proof that the fraction of orientations of K_n that are totally cyclic is at least $1 - \frac{2n+1}{2^{n-1}}$.

There are (at least) two ways to prove Proposition 7.28, and both of them boil down to the same idea. Given an orientation of a graph G, we can define an equivalence relation $\sim$ on $V(G)$, where $u \sim v$ if there is a directed path from u to v and a directed path from v to u. This divides the vertices of G into equivalence classes, called the *strongly connected components* of G. All of the edges between two strongly connected components must go in the same direction. For K_n, then, the edges that connect the strongly connected components provide a total order on the components. The number of orientations of K_n whose strongly connected components are each contained in a single block of the ordered set partition $\pi = (B_1, \ldots, B_k)$ is

$$2^{\binom{|B_1|}{2}+\cdots+\binom{|B_k|}{2}}.$$

Then you can apply inclusion-exclusion with the sets $A_{(B_1,B_2)}$ which consist of the orientations whose strongly connected components each lie either wholly in B_1 or wholly in B_2. Some care needs to be taken, though: A partition into four blocks can be written as the intersection of three different two-block sets, but it can also be written as the intersection of two different two-block sets. So some focused counting needs to be done with the coefficients in inclusion-exclusion.

The other approach leverages the stratification by number of blocks to define an exponential generating function and chug away.

Let's dismiss the details and focus on something rather miraculous. We can reinterpret the number $2^{\binom{|B_i|}{2}}$ as the number of simple graphs whose vertex set is B_i. Using that, we get

$$T_n = \sum_{H \subseteq E(K_n)} \sum_{\substack{\text{ordered set par-}\\ \text{titions } \sigma \text{ of the}\\ \text{components of } H}} (-1)^{\#\text{blocks}(\sigma)-1}.$$

The inner sum depends only on the number of components of H; denote that number by $f\big(\kappa(H)\big)$. More precisely, if we set

$$f(k) = \sum_{\sigma=(B_1,\ldots,B_r) \models [k]} (-1)^{r-1},$$

where the sum ranges over all ordered set partitions of $[k]$, then

$$T_n = \sum_{H \subseteq E(K_n)} f\big(\kappa(H)\big).$$

The first few values of f are

k	1	2	3	4
$f(k)$	1	-1	1	-1

which is highly suspicious. Indeed, we can prove that $f(k) = (-1)^{k-1}$ by induction.

Lemma 7.29. $f(k) = (-1)^{k-1}$ *for every* $k \in \mathbb{N}$.

Proof. Let $\mathcal{B}$ denote the set of ordered set partitions of $[k]$ that do not have $\{k\}$ as their final block. We first show that

$$\sum_{\sigma \in \mathcal{B}} (-1)^{\#\text{blocks}(\sigma)-1} = 0. \tag{$*$}$$

To do this, define the function $\varphi\colon \mathcal{B} \to \mathcal{B}$ as follows: If, in the ordered set partition $\sigma \in \mathcal{B}$, the element k is the only element in its block, then $\varphi(\sigma)$ is the ordered set partition that merges the block $\{k\}$ in σ with the block that follows it. If k is not the only element in its block, then $\varphi(\sigma)$ is the ordered set partition that removes k from its current block and inserts the block $\{k\}$ immediately before it. For example, if $k = 6$, then $\varphi(34, 126, 5) = (34, 6, 12, 5)$. The map φ is a sign-reversing involution on $\mathcal{B}$, which shows that the sum ($*$) vanishes.

So the sum in the definition of T_n reduces to a sum over the ordered set partitions of k in which the final block is $\{k\}$. These correspond exactly to the ordered set partitions of $[k-1]$. In fact, at this point, the only difference betweeen $f(k)$ and $f(k-1)$ is a sign: $f(k) = -f(k-1)$. Since $f(1) = 1$, that proves the claim. □

Plugging in this evaluation, we get the surprising result that

$$T_n = \sum_{H \subseteq E(K_n)} (-1)^{\kappa(H)-1}.$$

As you might expect, this can be generalized:

Theorem 7.30. *If G is any undirected graph (possibly with multiple edges or loops), the number of totally cyclic orientations of G is*

$$\sum_{H \subseteq E(G)} (-1)^{\kappa(H)-\kappa(G)}.$$

You can prove this, if you want, by following the same path that we used for $G = K_n$. There is a related formula for acyclic orientations:

Theorem 7.31. *If G is any undirected graph (possibly with multiple edges or loops), the number of acyclic orientations of G is*

$$\sum_{H \subseteq E(G)} (-1)^{|E(G)|-|V(G)|+\kappa(H)}.$$

Exercise 7.22. Prove Theorem 7.31 using Theorem 6.34 and Exercise 6.8.

7.8. the tutte polynomial

We've been building up to one of the most general algebraic structures for matroids: The Tutte polynomial. Just as with matroids themselves, there are many equivalent ways of formulating it, each with their own benefits. We'll define the Tutte polynomial by the so-called corank-nullity formula and prove the other formulations as theorems.

Definition 7.32. If $M = (E, \mathcal{B})$ is a matroid of rank d and $S \subseteq E$, the *corank* of S is $d - \operatorname{rank}(S)$; the *nullity* of S is $|S| - \operatorname{rank}(S)$.

Definition 7.33. There is a useful way of formulating corank and nullity so that they appear as dual concepts. Prove that the corank of S is the minimum size of a set T so that $S \cup T$ contains a basis of M; and prove that the nullity of S is the minimum size of a set T so that $S \setminus T$ is contained in a basis.

For graphical matroids, these quantities have expressions in terms of connected components. If $H \subseteq E(G)$, then in the matroid M_G,

$$\begin{aligned} \operatorname{corank}(H) &= \kappa(H) - \kappa(G) \\ \operatorname{nullity}(H) &= \kappa(H) + |E(H)| - |V(G)| \end{aligned}$$

Here, then, is the definition.

Definition 7.34. The *Tutte polynomial* of a matroid M on the ground set E is the bivariate polynomial

$$T_M(x,y) = \sum_{S \subseteq E} (x-1)^{\operatorname{corank}(S)} (y-1)^{\operatorname{nullity}(S)}$$

with the convention that $(x-1)^0 = (y-1)^0 = 1$. If $M = M_G$, then we write T_G for T_{M_G}.

One immediate reason for caring about the Tutte polynomial is that it is a joint generalization of the chromatic and flow polynomials.

Proposition 7.35. *For any graph G, we have*

$$\begin{aligned} \widetilde{\chi}_G(t) &= (-1)^{|V(G)| - \kappa(G)} T_G(1-t, 0) \\ C_G(t) &= (-1)^{|E(G)| - |V(G)| + \kappa(G)} T_G(0, 1-t). \end{aligned}$$

Exercise 7.23. Prove Proposition 7.35. (**Hint:** Use Theorem 6.34 for the chromatic polynomial.)

Beyond this, there are many specializations of the variables to integer values that give information on the graph. For example:

Proposition 7.36. *If G is a graph, then*

1. *$T_G(2,0)$ is the number of acyclic orientations of G.*
2. *$T_G(0,2)$ is the number of totally cyclic orientations of G.*
3. *$T_G(2,2) = 2^{|E(G)|}$.*
4. *$T_G(1,1)$ is the number of spanning forests of G.*

Proof. Parts 1 and 2 are Theorem 7.30, respectively. For part 3: After substituting $x = y = 2$ into Definition 7.34, each term becomes 1, so $T_G(2,2)$ is just the number of subgraphs of G, which is $2^{|E(G)|}$. For part 4, recall that we set $(x-1)^0 = (y-1)^0 = 1$ before we evaluate. So every term in the corank-nullity formula vanishes when we substitute $x = y = 1$ except those terms whose corank and nullity are both 0; each of these terms contributes 1 to the sum. These are exactly the bases of M_G, that is, the spanning forests of G. □

More generally, $T_M(1,1)$ is the number of bases of M.

Let's move on to properties of the Tutte polynomial more broadly. First up, it's multiplicative: $T_{M \sqcup N} = T_M T_N$. (We haven't defined the disjoint union of two matroids, but it's exactly what you expect: The bases of $M \sqcup N$ are the unions of a basis from M and a basis from N.) Also, the Tutte polynomial nicely displays matroid duality:

Proposition 7.37. $T_M(x,y) = T_{M^*}(y,x)$.

Proof. The corank of S in M is the nullity of $E \setminus S$ in M^*. □

More interestingly, it satisfies the deletion-contraction principle.

Definition 7.38. Suppose that $M = (E, \mathcal{B})$ is a matroid of rank d. If e is not a loop in M, then the *contraction* of M by e is the matroid M/e of rank $d-1$ on the ground set $E \setminus \{e\}$ whose bases are $\mathcal{B}/e := \{B \setminus \{e\} : B \in \mathcal{B} \text{ and } e \in B\}$. If e is not a coloop in M, then the *deletion* of e in M is the matroid $M \setminus e$ of rank d with ground set $E \setminus \{e\}$ and basis set $\mathcal{B} \setminus e := \{B \in \mathcal{B} : e \notin B\}$.

If e is a loop, then the set $\mathcal{B}/e$ is empty, and no matroid has an empty basis set. This is also why we require that e is not a coloop for deletion.

Theorem 7.39 (Deletion-contraction for Tutte polynomial). *The Tutte polynomial $T_M(x, y)$ is the unique bivariate polynomial parameter on matroids for which*

1. $T_M = T_{M/e} + T_{M \setminus e}$ *whenever e is neither a loop nor a coloop and*
2. $T_M = x^a y^b$ *if every element of M is a loop or a coloop, and M has a coloops and b loops.*

Note that condition 2 is equivalent to a list of three smaller conditions:

2a $T_M = yT_{M \setminus e}$ if e is a loop,
2b $T_M = xT_{M/e}$ if e is a coloop, and
2c $T_{\mathsf{Empty}} = 1$,

where Empty is the empty matroid: The matroid on the ground set $\emptyset$ whose collection of bases is $\{\emptyset\}$. These conditions explain why we used $x-1$ and $y-1$ in the definition of the Tutte polynomial instead of x and y; if we used the former, then the factors in conditions 2a and 2b would change to $x+1$ and $y+1$, respectively.

Proof sketch of Theorem 7.39. Condition 2c is immediate. To prove condition 2a, let $e \in E$ be a loop. If $e \in S$, then $\operatorname{corank}_M(S) = \operatorname{corank}_{M \setminus e}(S \setminus \{e\})$ and $\operatorname{nullity}_M(S) = \operatorname{nullity}_{M \setminus e}(S \setminus \{e\}) + 1$, so we can write

$$\begin{aligned} T_M(x,y) &= \sum_{\substack{S \subseteq E \\ e \notin S}} (x-1)^{\operatorname{corank}(S)}(y-1)^{\operatorname{nullity}(S)} + \sum_{\substack{S \subseteq E \\ e \in S}} (x-1)^{\operatorname{corank}(S)}(y-1)^{\operatorname{nullity}(S)} \\ &= T_{M \setminus e}(x,y) + (y-1)T_{M \setminus e}(x,y) \\ &= yT_{M \setminus e}(x,y). \end{aligned}$$

The proof of condition 2b is very similar.

To prove condition 1, take any element $e \in E$ that is neither a loop nor a coloop. For any subset $S \subseteq E$, we have

$$\left.\begin{aligned} \operatorname{corank}_M(S) &= \operatorname{corank}_{M \setminus e}(S) \\ \operatorname{nullity}_M(S) &= \operatorname{nullity}_{M \setminus e}(S) \end{aligned}\right\} \quad \text{if } e \notin S$$

and

$$\left.\begin{aligned} \operatorname{corank}_M(S) &= \operatorname{corank}_{M/e}(S \setminus \{e\}) \\ \operatorname{nullity}_M(S) &= \operatorname{nullity}_{M/e}(S \setminus \{e\}) \end{aligned}\right\} \quad \text{if } e \in S.$$

You can then use these to break up the sum as we did to prove condition 2a. The polynomial is determined by the initial conditions and the deletion-contraction recurrence, so it is unique. □

As a corollary, we find that the order in which deletion-contraction is applied doesn't matter. That's certainly not obvious. And that's the nice thing about having multiple characterizations: Some of them make certain properties of the Tutte polynomial obvious, even while others would make those same properties seem obscure.

Proposition 7.40. *The Tutte polynomial satisfies:*

- *If M and $\tilde{M}$ are isomorphic matroids, then $T_M = T_{\tilde{M}}$. (*Invariance*)*
- *$T_M(x, y) = T_{M^*}(y, x)$. (*Duality*)*
- *Every coefficient in $T_M(x, y)$ is positive. (*Positivity*)*

Proof. The corank-nullity formula makes invariance clear; deletion-contraction immediately implies positivity. And we already proved duality. □

There is one more common characterization of the Tutte polynomial; we turn to that next.

Definition 7.41. Let $M = (E, \mathcal{B})$ be a matroid and fix a total ordering $\preccurlyeq$ on E. Given a basis $I \in \mathcal{B}$, an element $e \in I$ is called *internally active* if there is no $e' \preccurlyeq e$ such that $I \setminus \{e\} \cup \{e'\}$ is a basis; an element $f \in E \setminus I$ is called *externally active* if there is no $f' \preccurlyeq f$ such that $I \cup \{f\} \setminus \{f'\}$ is a basis. The number of interally and externally active elements for I is denoted $\operatorname{int}(I)$ and $\operatorname{ext}(I)$, respectively.

The definition might be a bit clearer in the graphical case. In the graphical matroid M_G, an element $e \in I$ is internally active if there is a partition of $V(G)$, say $A \sqcup B$, so that e is the minimal element in the collection of edges connecting A and B. (In jargon: e is the minimal edge of a cut-set.) And $f \in E(G) \setminus I$ is externally active if it is the minimal edge in the unique cycle in $I \cup \{f\}$.

Theorem 7.42. *Given a matroid M with basis set $\mathcal{B}$ and any total order on its ground set,*

$$T_M(x, y) = \sum_{B \in \mathcal{B}} x^{\operatorname{int}(B)} y^{\operatorname{ext}(B)}.$$

Proof. By induction. The base case, when M contains only loops and coloops, is not so hard: M has a single basis B consisting of all the coloops. Every element of $E \setminus B$ is externally active, and every element of B is internally active.

Now suppose that M has some positive number of elements that are neither loops nor coloops; we induct on this number. If we take e to be the maximal element that is neither a loop nor coloop under the total order, then

$$\left.\begin{aligned} \operatorname{int}_M(B) &= \operatorname{int}_{M \setminus e}(B) \\ \operatorname{ext}_M(B) &= \operatorname{ext}_{M \setminus e}(B) \end{aligned}\right\} \quad \text{if } e \notin B$$

and

$$\left.\begin{aligned} \operatorname{int}_M(B) &= \operatorname{int}_{M/e}(B \setminus \{e\}) \\ \operatorname{ext}_M(B) &= \operatorname{ext}_{M/e}(B \setminus \{e\}) \end{aligned}\right\} \quad \text{if } e \in B.$$

As we did for deletion-contraction, use this to decompose the sum; then use the induction hypothesis. $\square$

7.9. polymatroids

Matroids are pretty abstract. But some people thought they weren't abstract enough, whence polymatroids. What matroids are to graphs, polymatroids are to hypergraphs. And while a matroid is realized as a collection of vectors, a polymatroid is realized as a collection of subspaces of a field. At a high level, polymatroids allow dimensions to mix.

Like matroids, polymatroids can be defined in a constellation of equivalent ways. Here's one:

Definition 7.43. A function $f\colon \mathcal{P}([n]) \to \mathbb{R}$ is called *submodular* if $f(I)+f(J) \geq f(I \cup J)+f(I \cap J)$ for every $I, J \subseteq [n]$. The range of an *integer submodular function* lies in $\mathbb{Z}$.

A *polymatroid* is a submodular function ρ that is nonnegative and non-decreasing: if $I \subseteq J$, then $\rho(I) \leq \rho(J)$. If f is submodular, then $f + c$, for any $c \in \mathbb{R}$, is submodular, so we can always ensure that $f \geq 0$ and $f(\emptyset) = 0$. We can also transform an arbitrary nonnegative submodular function to make it monotone.

Exercise 7.24. Let $w\colon [n] \to \mathbb{R}_{\geq 0}$ be an arbitrary function and define $f\colon \mathcal{P}([n]) \to \mathbb{R}_{\geq 0}$ by $f(I) = \sum_{i \in I} w(i)$. Show that f is a monotone submodular function. Also show that if w and $\tilde{w}$ are two weight functions that give rise to the same monotone submodular function, then $w = \tilde{w}$.

Some polymatroids are realizable as a collection of subspaces of a vector space. A *subspace arrangement* in F^n is simply a collection of subspaces in the vector space F^n. The polymatroid associated to a subspace arrangement $\mathcal{V} = \{V_1, \ldots, V_n\}$ is

$$\rho(I) = \dim\Big(\operatorname{span} \bigcup_{i \in I} V_i\Big).$$

A polymatroid that arises in this way is called a *realizable* polymatroid.

Next, we'll introduce two different characterizations of polymatroids.

Definition 7.44. A *generalized permutohedron* is a convex polytope $P \subseteq \mathbb{R}^n$ for which each edge is parallel to a vector of the form $e_i - e_j$ with $i \neq j$. If every vertex of P is in $\mathbb{Z}^n$, then P is called an *integer generalized permutohedron*.

Exercise 7.25. Prove that every permutohedron is a generalized permutohedron.

Definition 7.45. A set $S \subseteq \{\mathbf{x} \in \mathbb{Z}^n : x_1 + \cdots + x_n = k\}$ is called *M-convex set* if, for every $\mathbf{x}, \mathbf{y} \in S$ and $i \in [n]$ with $x_i > y_i$, there is a $j \in [n]$ with $x_j < y_j$ such that $\mathbf{a} - e_i + e_j$ and $\mathbf{b} + e_i - e_j$ are both in S.

M-convexity is some notion of convexity for the discrete space $\mathbb{Z}^n$: If $\mathbf{a}$ and $\mathbf{b}$ are in S, then you can move them closer to each other inside the set S.

Proposition 7.46. *Integer submodular functions, integer generalized permutohedra, and M-convex sets are cryptomorphic (that is, they encode the same information.*

To prove this, we just need to show bijections between the various objects. There are some fairly simple ones. To get from a submodular function $f\colon \mathcal{P}([n]) \to \mathbb{Z}$ to a generalized permutohedron, set

$$P_f = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i \in I} x_i \leq f(I) \text{ for all } I \subseteq [n] \text{ and } \sum_{i=1}^{n} x_i = f([n])\}.$$

The inverse of the map $f \mapsto P_f$ sends P to the submodular function $f_P\colon \mathcal{P}([n]) \to \mathbb{Z}$ given by

$$f_P(I) = \max_{\mathbf{x} \in P} \sum_{i \in I} x_i.$$

Exercise 7.26. Verify that these are inverse maps. (**Hint**: The key point to verify is this: If f is a submodular function, each of the hyperplanes $\sum_{i \in I} x_i \leq f(I)$ is tangent to P at some point.)

To get from an integer generalized permutohedron $P \subseteq \mathbb{R}^n$ to an M-convex set S_P, simply take $S_P = P \cap \mathbb{Z}^n$. The inverse map sends S to $P_S = \operatorname{conv}(S)$. Because all the extremal points of P lie in the integer lattice, these maps are inverses.

Interpreting matroids (the regular kind) as polytopes makes duality super easy: $P_{M^*} = \mathbf{1} - P_M$, where $\mathbf{1}$ is the all-ones vector. So duality of matroids amounts to reflection about the point $\frac{1}{2}\mathbf{1}$.

We'll close this section by describing one new class of generalized permutohedra that derives from graphs.

Definition 7.47. A *zonotope* is a [Minkowski sum] of a finite number of line segments. Given a graph G on the vertex set $[n]$, the *graphical zonotope* associated to G is

$$Z_G := \sum_{\substack{ij \in E \\ i<j}} [e_i, e_j],$$

where Σ denotes the Minkowski sum.

Up to translation, $Z_G = \sum_{ij} [0, e_j - e_i]$.

Actually, the face poset of Z_G is dual to the face poset of the hyperplane arrangement $\mathcal{A}_G$; the transformation taking Z_G to $\mathcal{A}_G$ sends each line segment in Z_G to the central hyperplane with that line segment as a normal vector. Zonotopes, then, are very similar to what we've already studied, but they hold a bit more information; for example, you can take their volume.

There is a notion of a *cographical zonotope* that exists in the flow space of G. (The *flow space* of a graph is the set of all flows $f\colon E(G) \to$ R.) It is a $(|E(G)| - |V(G)| + 1)$-dimensional vector space. In the dual of the flow space, there are the evaluation vectors $v_e\colon f \mapsto f(e)$ for each $e \in E(G)$; the cographical zonotope is

$$Z_G := \sum_{e \in E(G)} [0, v_e]$$

in the dual flow space.

There are interpretations of various quantitative information derived from Z_G and Z_G:

	Z_G	Z_G
# vertices	acyclic orientations	totally cyclic orientations
# lattice points	subforests	spanning subgraphs
volume	spanning trees	spanning trees.

Some of them follow directly from the bijection between graphical zonotopes and graphical hyperplane arrangements; others don't. You can try to prove them, if you want.

A. appendix

A.1. finite fields

Throughout this section, F is a finite field with q elements.

Proposition A.1. *q is a power of a prime.*

Proof. Let p be the *characteristic* of F; that is, the least natural number such that

$$\underbrace{1+\cdots+1}_{p}=0.$$

Such a number exists, since it's just the order of the element 1 in the additive group of F. We claim that p must be prime: If $p=ab$ with $a,b>0$, then

$$(\underbrace{1+\cdots+1}_{a})((\underbrace{1+\cdots+1}_{b})=0,$$

so one of the terms on the left must be zero. Since we chose p to be minimal, either $a=p$ or $b=p$. So p is prime.

This means that F is a vector space over Z$/p$Z, where

$$a\cdot x=\underbrace{x+\cdots+x}_{a}$$

for every $a\in$ Z$/p$Z and $x\in F$. (You can check that all of the axioms of a vector space are satisfied.) In particular, F has a basis as a vector space, say $e_1,\ldots,e_n$. So every element of F can be uniquely expressed as a linear combination

$$a_1e_1+\cdots a_ne_n$$

with $a_i\in$ Z$/p$Z. There are exactly p^n such linear combinations, so F has exactly p^n elements. □

That's a neat little argument. So there's no field with exactly 28 elements, for example. It doesn't, however, show that there *is* a field with p^n elements. We do that next.

Proposition A.2. *There is a finite field with p^n elements for every prime p and positive integer n.*

It uses one lemma that relies on a bit of commutative algebra. If you've not yet encountered any ring theory, then it's probably not worth reading the rest of this section.

Lemma A.3. *If k is a field and f is an irreducible polynomial in $k[x]$, then the ideal (f) is a maximal ideal in $k[x]$.*

The proof of this just relies on the fact that in the ring $k[x]$, any two polynomials have a greatest common divisor. Anyway, let's skip to the good stu.

Proof of Proposition A.2. Suppose we can find some irreducible polynomial $f\in$ (Z$/p$Z)$[x]$ with $\deg(f)=n$. Then $F:=$ (Z$/p$Z)$[x]/(f)$ is a field, since (f) is a maximal ideal. It's not too hard to prove that every element of F is representable in the form

$$a_0+a_1x+\cdots+a_{n\ \ 1}x^{n\ \ 1}+(f),$$

and that every element in F is *uniquely* representable in this way. This means that F has exactly p^n elements.

So we just need to find an irreducible polynomial with degree n. It turns out to be hard to find an explicit one, but proving that one exists just comes down to a bit of counting. If g is a monic polynomial in (Z$/p$Z)$[x]$ with degree n, then it can be factored as the product of two monic polynomials $h_1 \cdot h_2$ with $\deg(h_1) + \deg(h_2) = n$ and $\deg(h_1), \deg(h_2) \geq 1$. How many monic polynomials of degree k are there? Easy: $p^{k\ \ 1}$. We may assume that $\deg(h_1) \leq \deg(h_2)$, so the number of *reducible* polynomials of degree n is at most

$$\sum_{k=1}^{bp/2c} p^{k\ \ 1} p^{n\ \ k\ \ 1} \leq \frac{p}{2} p^{n\ \ 2} = \frac{1}{2} p^{n\ \ 1}.$$

Since there are $p^{n\ \ 1}$ monic polynomials of degree n, some of them must be irreducible. □

Proving that two finite fields with the same number of elements are always isomorphic is more di□cult. For a proof of this and other facts about finite fields, see Keith Conrad's notes.

A.2. systems of linear equations

When does a system of m linear equations

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= c_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= c_2 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= c_n \end{aligned}$$

have a common solution $x = (x_1, \ldots, x_n)$ in the vector space F^n? If we make the coe□cients into a matrix $A = (a_{i,j})$ and the right-hand numbers into a vector $c = (c_1, \ldots, c_n$, then we're looking for a solution to the equation $Ax = c$. This is simple if c is the zero vector $\mathbf{0}$; then $\mathbf{0}$ is itself a solution. Actually, x is a solution if and only if it is orthogonal to every row of A. Recall that the *rank* of a matrix A is the dimension of the subspace of F^n spanned by the rows of A. Therefore $Ax = \mathbf{0}$ if and only if x is in the orthogonal subspace to the row space of A. In particular:

Lemma A.4. *The set of solutions to the equation $Ax = \mathbf{0}$, where A is an $m \times n$ matrix, is a subspace of F^n and has dimension $n - \operatorname{rank}(A)$.*

What if $c \neq \mathbf{0}$? Then we can form the *augmented matrix* $\bar{A} = [A|c]$ (obtained by appending c as an extra column). A solution to $Ax = c$ exactly corresponds to a solution $\bar{A}y = \mathbf{0}$ with $y_{n+1} = -1$, and vice versa. This almost reduces it to the previous problem. This is the key.

Proposition A.5. *There is a solution to the equation $Ax = c$ if and only if $\operatorname{rank}(A) = \operatorname{rank}(A|c)$.*

Proof. We use V to denote the solution space to $Ax = \mathbf{0}$ and W to denote the solution space to $\bar{A}y = \mathbf{0}$. By the lemma, the dimension of W is $n + 1 - \operatorname{rank}(\bar{A})$. If A has a collection of k row vectors that are linearly independent, then the same row vectors in $\bar{A}$ are linearly independent, as well; so $\operatorname{rank}(\bar{A}) \geq \operatorname{rank}(A)$.

Since $(x_1, \ldots, x_n, 0) \in W$ whenever $(x_1, \ldots, x_n) \in V$, we have $W \supseteq V \oplus \{0\}$. If $\operatorname{rank}(\bar{A}) > \operatorname{rank}(A)$, then $\dim(W) = n+1-\operatorname{rank}(\bar{A}) \leq n - \operatorname{rank}(A) = \dim(V)$. But W contains $V \oplus \{0\}$; this means that $W = V \oplus \{0\}$. In particular, the last coordinate of every vector in W is 0, not -1; so there is no solution to $Ax = c$.

On the other hand, if $\operatorname{rank}(A) = \operatorname{rank}(\bar{A})$, then $\dim W = \dim V + 1$, so W contains a vector y with $y_{n+1} \neq 0$. Then $z = -\frac{1}{y_{n+1}} y$ is a vector in W whose last coordinate is -1; the vector $x = (z_1, \ldots, z_n)$ is a solution to the equation $Ax = c$. □

We need a few more fact about matrices. Recall that a *minor* of a matrix A is the determinant of a square submatrix of A.

Lemma A.6. $\ker A = (\operatorname{im} A^T)^?$.

Proof. A vector $x \in \mathrm{R}^n$ is in the kernel of A if and only if $\langle Ax, y\rangle = 0$ for every $y \in \mathrm{R}^m$. From $\langle Ax, y\rangle = \langle x, A^T y\rangle$, the claim follows. □

Lemma A.7. *The rank of A is also the rank of A^T.*

Proof. Take the previous lemma and calculate dimensions using the rank-nullity theorem:

$$n - \mathrm{rank}_c(A) = \dim(\ker A) = n - \dim(\operatorname{im} A^T) = n - \mathrm{rank}_c(A^T),$$

where $\mathrm{rank}_c(A)$ denotes the column rank of A (the dimension of the image of A; alternatively, the dimension of the span of the column vectors of A). Therefore $\mathrm{rank}_c(A) = \mathrm{rank}_c(A^T)$. Substituting $\mathrm{rank}_c(A) = \mathrm{rank}(A^T)$ finishes the proof. □

Now meet one final participant in our parade of lemmas.

Lemma A.8. *The rank of A is at least r if and only if every $r \times r$ minor of A is nonzero.*

Proof. If $\mathrm{rank}(A) < r$, then any choice of r row vectors of A are linearly dependent. So the determinant of the projection of these r vectors into any subspace is 0.

If $\mathrm{rank}(A) \geq r$, then there are r linearly independent row vectors; let B be the submatrix of A that consists of these rows. By Lemma A.7, the matrix B has r linearly independent column vectors. The corresponding $r \times r$ matrix is a submatrix of A with nonzero determinant. □

Index